

2 ejemplares
N. 13

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO
FACULTAD DE CIENCIAS



LA TRANSFORMACION DE VARIABLES

99
XCI

T E S I S
PARA OBTENER EL TITULO DE
A C T U A R I O
P R E S E N T A

RICARDO ARTURO LORDA ANDRADE

México, D. F.

1979

10013



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

	pág.
Introducción	1
Capítulo 1	
El uso de las transformaciones	4
Capítulo 2	
Métodos de Transformación	
2.1 Transformación de la variable dependiente	13
2.2 Transformación de la variable independiente	28
2.3 Transformación de ambas variables	45
Capítulo 3	
Transformaciones específicas que más se han utilizado en la práctica	
3.1 Transformación Potencia	58
3.2 Transformación Recíproca	59
3.3 Transformación de Raíz Cuadrada	60
3.4 Transformación de la Raíz Cúbica	63
3.5 Transformaciones logarítmica e hiperbólica	63
3.6 Transformación angular	67
3.7 Transformación logarítmica doble	68
3.8 Transformación Integral de Probabilidad	69
3.9 Transformación coordinada	69
3.10 Transformación Probit	70
Capítulo 4	
Procedimientos para verificar la utilidad de la transformación	
4.1 Obtención del sesgo de la transformación	72
4.2 Utilización del coeficiente de correlación múltiple, R^2 , para verificar la transformación	77
4.3 Un procedimiento para averiguar los efectos de la transformación	83

Capítulo 5

Conclusiones	89
Apéndice	92
Bibliografía	96

I N T R O D U C C I O N

En la mayoría de los libros introductorios de Estadística, y en particular de Análisis de Regresión, se hace mención de la transformación de variables como uno de los métodos para conseguir mejorar -de acuerdo a determinados criterios- o simplificar el modelo estadístico en estudio. Sin embargo, en la mayoría de ellos se reducen a mencionar muy someramente algunos de los métodos de transformación de variables o, lo que es más común, se reducen a dar algunos ejemplos de transformaciones y la utilidad que pueden tener en términos muy generales. Si bien es cierto que existen algunos otros libros más especializados que tratan más a fondo el problema de las transformaciones, también lo es que éstos no son, por una parte, muy asequibles -al menos actualmente- y, por otra, son demasiado extensos para propósitos prácticos.

Algunos trabajos se limitan a analizar a fondo alguna transformación o método de transformaciones con el fin de dar una ayuda a aquellos que más la utilizan en la práctica y que no están plenamente familiarizados con el método estadístico; o hacen énfasis en el desarrollo y alcances de tal método, con un enfoque fundamentalmente teórico.

Pensamos, pues, que se ha quedado un poco olvidado uno de los aspectos prácticos importantes: las pretensiones y el alcance tanto de cada uno de los métodos como de las transformaciones par-

ticulares que se han estudiado, así como los procedimientos para verificar los objetivos conseguidos por la transformación aplicada.

Con el presente trabajo pretendemos analizar primero la necesidad, o mejor, la conveniencia del uso de las transformaciones (Capítulo 1); presentamos también una recopilación de los métodos de transformación más conocidos de la forma más breve posible pero tratando de que queden claramente explicados y asequibles (Capítulo 2). Los ejemplos que damos tratan de ser una ayuda para conseguir este fin. No constituye nuestro propósito, como veremos más adelante, el hacer un desarrollo de cada uno de los métodos ni el fijar nuestra atención en los aspectos algebraicos o de cálculo de ellos mismos; este trabajo ya se ha hecho con más o menos amplitud y precisión (si se desea profundizar en este aspecto, se puede ver p. ej. Guerrero y Hernández, 1974, o cada uno de los artículos originales que presentan el método en cuestión). Describimos primero los métodos para transformar la variable dependiente, luego aquéllos utilizados para la variable independiente y, por último, un procedimiento que puede usarse para ambas.

En el tercer capítulo presentamos una recopilación de las transformaciones más utilizadas, haciendo énfasis en el uso que se les ha dado y en lo que se consigue con ellas. Exponemos también algunas que sirven para transformar poblaciones con una distribución específica distinta de la Normal.

En el Capítulo 4 describimos algunos procedimientos, bastante recientes, para verificar la utilidad de las transformaciones. Por último, en el apéndice damos unas tablas necesarias para la utilización de uno de los procedimientos del Capítulo 2 y la definición de las funciones enteras utilizadas en el primer método del Capítulo 4.

Con este trabajo no intentamos, por supuesto, hacer un estudio exhaustivo sobre el tema, es obvio que las limitaciones de espacio, tiempo y de la literatura accesible nos lo impiden.

C A P I T U L O 1

EL USO DE LAS TRANSFORMACIONES

Tanto las transformaciones como los métodos de transformación que se han estudiado, han surgido en su gran mayoría al intentar resolver problemas de Análisis de Regresión y es ahí, por tanto, donde más se han utilizado. Tomando en cuenta esto, en adelante nos reduciremos prácticamente a referirnos a los modelos de regresión, aunque habrá que tener presente que pueden perfectamente aplicarse a modelos de otro tipo (p. ej. en los modelos de diseño de experimentos).

Un modelo de regresión es un modelo estadístico de la forma:

$$y = f(x_1, x_2, \dots, x_k; \theta_1, \theta_2, \dots, \theta_p) + \epsilon$$

donde y ; x_1, x_2, \dots, x_k y ϵ son variables aleatorias y $\theta_1, \theta_2, \dots, \theta_p$ parámetros desconocidos -por lo general- que intervienen en el modelo. En general, prácticamente -y es el caso más común- sólo se conocen algunas, digamos $p-1$, de las K variables independientes, x_1, x_2, \dots, x_k ; se cuenta con un conjunto de n observaciones de cada una de las $p-1$ variables independientes y de la variable dependiente y . Como la relación funcional f es desconocida, se pretende encontrar una relación estimada en base a las observaciones.

Dentro de los modelos de regresión, se ha visto que los modelos lineales tienen una gran utilidad práctica pues muchas situaciones

ciones pueden representarse adecuadamente haciendo uso de ellos y además están sustentadas en bases teóricas ampliamente desarrolladas. Aunque en este trabajo no nos limitaremos a este tipo de modelos, pensamos que es importante definirlos. Un modelo lineal de regresión es un modelo que puede ser escrito en la forma:

$$\underline{y} = \underline{x} \underline{\beta} + \underline{\epsilon}$$

donde \underline{y} es el vector de observaciones y \underline{x} una matriz de forma conocida, de dimensiones $(n \times 1)$ y $(n \times p)$, respectivamente. $\underline{\beta}$ es el vector de los parámetros, de dimensión $(p \times 1)$ y $\underline{\epsilon}$, de $(n \times 1)$, el vector de "errores".

Podemos decir que el propósito inicial del análisis de regresión es aproximar la relación funcional que liga a las variables independientes con la variable dependiente ya que en la mayoría de los casos aquella se desconoce o, si no es así, es una relación demasiado complicada para ser comprensible o, al menos, para describirla en términos simples. Esta aproximación se hace, por lo general, con uno de los dos fines siguientes:

- 1) Apreciar los efectos que las variables independientes consideradas producen en la variable dependiente, o
- 2) Poder predecir o pronosticar el valor de la variable dependiente para ciertos valores dados de las variables independientes.

En el primer caso se desea analizar precisamente el efecto

producido por ciertas variables. Hemos de trabajar con las observaciones tal cual, pues es lo que interesa. No tiene sentido, pues, en este tipo de circunstancias, la transformación de variables porque estaríamos alterando la situación original. Supongamos p. ej. que deseamos medir el efecto que tiene la lluvia x , medida en m.m., sobre cierta variable, y . Si transformamos la variable x a $\log x^2$, p. ej., puede ser que efectivamente el $\log x^2$ tenga gran influencia en los valores de y , pero hemos perdido nuestro objetivo: ¡no hemos logrado saber si x -la lluvia- produce algún efecto en y !, pues físicamente no tenemos una explicación para $\log x^2$.

En cambio, en el segundo caso, estamos interesados en la predicción del valor de una variable, sin importar la forma concreta en que se haga ésta. Resulta claro entonces que es en este tipo de situaciones cuando puede tener sentido el uso de las transformaciones. Puede suceder pues que las variables transformadas no tengan ningún significado en la realidad, pero eso no importa ahora mientras sirvan para predecir la variable que nos interesa.

Una vez que ha quedado claro cuándo puede tener sentido el transformar una variable, sería lógico preguntarse qué pretendemos al hacerlo, ¿por qué razones es útil transformar los datos?, ¿no sería lo ideal trabajar con los datos originales? Efectivamente, lo ideal sería que con los datos originales pudiéramos obtener un modelo que sea sencillo, que represente lo mejor posible la si-

tuación real que se presenta y que sea útil para nuestros propósitos; pero no cabe duda que esto es bastante difícil en muchos casos.

Una razón básica para transformar los datos es intentar hacer el análisis más simple posible entre todos los que sean factibles. Por ejemplo, los datos originales pueden requerir el uso de un modelo de regresión de 2º orden, mientras que, después de una transformación, un modelo de primer orden puede ser perfectamente adecuado y tal simplificación llevará consigo que las implicaciones de los datos puedan ser entendidas más fácilmente. Incluso en el caso de que pueda usarse una relación de primer grado para los datos originales, puede suceder que con una transformación se extienda el rango de las variables sobre el cual es válida la relación de primer orden, lo que nos permitiría tener un rango de predicciones (interpolaciones, en este caso) más amplio.

Originalmente las transformaciones se utilizaron más que nada para hacer más válido el análisis de la varianza (ver p. ej. Bartlett (1947), p. 39 o Curtiss (1943), p. 107), ya que éste incluye la suposición de tener una varianza constante en los errores (residuales) y si ésta tiende a cambiar con la media de las observaciones, sólo podrá ser estabilizada por medio de un adecuado cambio de escala. Sin embargo, no es solamente el conseguir una varianza constante lo que se pretende. En general, podemos decir

que las técnicas usuales para el análisis de los modelos lineales han sido desarrolladas con los siguientes supuestos básicos:

- a) Varianza constante en el error
- b) Aditividad en los efectos
- c) Las observaciones tienen una distribución Normal, y
- d) Las observaciones son independientes.

La primera hipótesis se hace usualmente porque simplifica las técnicas de estimación; si esta hipótesis se cumple, los estimadores de mínimos cuadrados son también estimadores lineales insesgados de varianza mínima. Sin esta suposición, un análisis de mínimos cuadrados con pesos, nos proporciona estos estimadores (ver p. ej. Draper y Smith (1966), pp. 77-81).

La hipótesis de aditividad es importante en la interpretación de los datos; no es una hipótesis que sea siempre necesaria para la estimación o para las pruebas de hipótesis aunque en muchas circunstancias será necesaria para asegurar la identificación de los parámetros.

El tercer supuesto es sumamente importante en las pruebas de hipótesis, pues la Normalidad en las observaciones nos conduce a procedimientos estándar de prueba relativamente simples que han sido investigados a fondo y a distribuciones conocidas.

Por último, la hipótesis de independencia es también muy útil pues nos permite obtener estimadores de la varianza de los pa-

rámetros así como hacer pruebas de hipótesis en base a distribuciones conocidas, lo cual, a su vez, permite hacer predicciones válidas.

Si éstas hipótesis no se cumplen se puede proceder de varias formas, Graybill (1961, p. 332) menciona 4 alternativas (que en realidad se reducen a tres):

- 1) Usar procedimientos no paramétricos que son válidos para supuestos muy generales o, en general, intentar otros métodos de análisis con supuestos que se ajusten a los datos originales, al menos de mejor forma (en ocasiones esto supondrá el desarrollo de nuevos métodos).
- 2) Ignorar que las hipótesis no se cumplen y proceder como si lo hicieran.
- 3) Transformar las variables observadas para lograr que cumplan las condiciones, al menos de una forma aproximada; y
- 4) Si la prueba es insensitiva a las condiciones pedidas, se dice que es robusta. Si una prueba es robusta, el procedimiento 2) es útil.

Como menciona Turkey (1957, pp. 602, 609), si es posible encontrar una transformación satisfactoria, la alternativa 3) será casi siempre más simple que el intentar otros métodos, sobre todo

cuando haya que desarrollar un método nuevo y, obviamente, más atractiva que la alternativa 2) excepto en el caso de que una prueba sea robusta.

Podemos afirmar, de esta manera, que el propósito de toda transformación es incrementar el grado de aproximación a las hipótesis hechas en los métodos convencionales de análisis. Hasta la fecha, como veremos posteriormente, los métodos de transformaciones que se han desarrollado, han fijado su atención en cumplir fundamentalmente las primeras tres hipótesis mencionadas (o al menos una de ellas): varianza constante, aditividad y Normalidad en las observaciones. Quizá esto se deba, como afirman Box y Tidwell (1962, p. 531) a que frecuentemente la suposición de independencia no está en cuestión o a que se ha introducido una aleatorización de manera que esto permite hacer el análisis como si las observaciones fueran independientes.

Box y Cox (1964, pp. 211-4), sin embargo, afirman que en cualquier caso estamos interesados no exactamente en encontrar una transformación que justifique las hipótesis sino más bien en encontrar -si es posible- una métrica en cuyos términos los resultados puedan expresarse de una forma más concisa.

Si una transformación se elige con el fin de cumplir una de las hipótesis, es obvio que no podemos esperar que se satisfagan todas, aunque sabemos que se encuentran relacionadas. Las trans-

formaciones elegidas para satisfacer la condición de varianza constante, con frecuencia tienen el efecto de mejorar la proximidad de la distribución a la distribución Normal. Una correlación de la varianza con el nivel medio en la escala original implica frecuentemente una excesiva desviación que tiende a ser eliminada después de una transformación adecuada. La validez de cualquier supuesto de Normalidad debe ser verificada, ya que mientras se sabe que las desviaciones moderadas de la distribución Normal no son serias, cualquier desviación grande es probable que afecte la validez de las pruebas de significancia.

Como ya se mencionó, algunas veces se hace la aseveración de que la escala original es la más relevante para hacer todas las estimaciones, así como la más comprensible, lo cual constituye un argumento en contra de las transformaciones sin un buen fundamento. Sin embargo, este argumento pierde fuerza al recordar que si la varianza de los datos cambia con el nivel medio para diferentes bloques o grupos, un promedio de los valores de la variable dependiente para el modelo observado no es necesariamente el mejor estimador del valor de la variable dependiente para el modelo correcto, y el promedio en la escala transformada será frecuentemente el mejor estimador cuando re-convertimos a la escala original.

Es más sencillo y más efectivo trabajar con variables para las cuales el modelo es lineal y aditivo. No siempre es posible

elegir una transformación para cumplir la condición de varianza constante y que simultáneamente sea más razonable para la hipótesis de aditividad, aunque puede suceder que la selección hecha con este criterio aumente el grado de proximidad a un modelo aditivo. En algunos casos en que se cuenta con información suficiente -procedente de la naturaleza de los datos- acerca de la variabilidad, podemos decidir abandonar las ventajas de que se dé la hipótesis de varianza constante y seleccionar la transformación para lograr únicamente la aditividad, dando pesos apropiados a las observaciones dependiendo de la variabilidad conocida.

Como resulta obvio, las transformaciones han de aplicarse una vez que se han verificado las hipótesis del modelo inicial propuesto haciendo uso de los métodos conocidos para este fin, lo contrario no tendría sentido. Además, antes de aplicar alguna transformación, habrá que analizar:

- 1º Si tiene sentido el uso de transformaciones
- 2º Si puede resultar conveniente
- 3º Cuáles son los objetivos de la transformación por hacer, y
- 4º Qué método se va a utilizar para ver si cumple los objetivos.

C A P I T U L O 2

METODOS DE TRANSFORMACION

Presentamos aquí los principales métodos de transformación que se han desarrollado hasta la fecha. Hemos creído conveniente dividirlos en tres apartados: métodos para transformar la variable dependiente, para la variable independiente y, por último, un procedimiento que permite transformar cualquiera de las dos. A continuación de cada método -que se expone lo más brevemente posible, por las razones ya mencionadas- se da un ejemplo con el fin de hacerlo más comprensible. La mayoría de los ejemplos están tomados del artículo original que desarrolla el método.

2.1 TRANSFORMACION DE LA VARIABLE DEPENDIENTE

METODO DE BARTLETT (1947)

Parece ser que antes del artículo escrito por Bartlett (1947) sólo se había estudiado la aplicación de algunas transformaciones específicas (como la transformación de la raíz cuadrada, la logarítmica, la inversa del seno, etc.), de forma que podemos decir que es éste el primer intento de un método formal de transformación.

El autor establece un método con una aplicación particular al análisis de la varianza. Entre las condiciones requeridas para conseguir un análisis preciso, cuando se hace de la manera usual,

está la de tener una varianza constante en los errores. Este es pues el propósito fundamental del método: conseguir estabilizar la varianza y , en particular, cuando ésta tiende a cambiar con el nivel medio de las observaciones. Cuando la media y la varianza están relacionadas es usual no tener Normalidad, de manera que podemos afirmar que el método pretende también aproximarse en cierto grado a la distribución Normal.

El método es el siguiente:

Supongamos que la $E(y)=m$ y que $\text{Var}(y)=f(m)$, es decir que la esperanza y la varianza están relacionadas mediante la función f^1 . Se pretende encontrar una función g tal que $E(g(y))$ no esté relacionada con la $\text{Var}(g(y))$ y además que $V(g(y))=c^2$, donde c es una constante.

Si $g(y)$ es diferenciable de orden n , podemos expandirla en series de Taylor alrededor de m , de manera que si despreciamos los términos de grado superior a uno, tendremos:

$$g(y) \approx g(m) + g'(m) (y-m)$$

por tanto,

$$E(g(y)) \approx E(g(m) + g'(m) (y-m)) = g(m)$$

y,

$$\begin{aligned} V(g(y)) &= E(g(y) - g(m))^2 \approx E(g'(m) (y-m))^2 \\ &= g'(m)^2 V(y) = g'(m)^2 f(m) \end{aligned}$$

1 En general podríamos decir que la esperanza y la varianza de una misma variable están relacionadas, al menos en cuanto a la escala (no tendría sentido p.ej. que una se diera en micras y la otra en km.). Sin embargo, aquí nos referimos a una relación funcional.

Entonces, si deseamos que $V(g(y))$ sea constante, digamos

$E(g(y)) = c^2$, tendremos:

$$g'(m)^2 f(m) = c^2, \quad g'(m) \sqrt{f(m)} = c$$

$$g'(m) = c / \sqrt{f(m)}$$

y, entonces:
$$g(m) = \int \frac{c \, dm}{\sqrt{f(m)}}$$

Ejemplos:

El primer ejemplo está tomado del artículo original.

- 1) Supongamos que la desviación estándar tiende a ser proporcional a la media, m .

Tendremos: $f(m) = k^2 m^2$, k constante

$$g(m) = \int \frac{c \, dm}{\sqrt{k^2 m^2}} = \int \frac{c \, dm}{k m} = \frac{c}{k} \ln m$$

Es decir, $g(m)$ sería proporcional al $\ln m$. Esto indicaría entonces el uso de la transformación $g(y) = \ln y$.

- 2) Si y se distribuye binomial, entonces $E(y) = np$, $V(y) = npq$.

Entonces si $m=np$, $f(m) = \frac{mn-m^2}{n}$, por tanto:

$$g(m) = \int \frac{c \, dm}{\sqrt{\frac{mn-m^2}{n}}} = c \sqrt{n} \int \frac{dm}{\sqrt{mn-m^2}} = c \sqrt{n} \arcsin\left(\frac{2m-n}{n}\right)$$

Es decir, hemos de usar la transformación inversa del seno.

METODO DE BOX Y COX (1964).

El procedimiento supone que después de una transformación, adecuada de y a $y^{(\lambda)}$, donde $y^{(\lambda)}$ es la variable transformada mediante el parámetro λ , se tendrá que:

- a) El valor esperado de las observaciones transformadas estará descrito por un modelo de una estructura simple;
- b) La varianza del error será constante; y
- c) Las observaciones se distribuirán Normalmente.

Demuestran que la función de verosimilitud de $y^{(\lambda)}$, maximizada de λ, y -en un enfoque bayesiano- la aproximada distribución marginal de λ , son ambas proporcionales a una potencia negativa de la suma de cuadrados de los residuales para la variable $z^{(\lambda)} = y^{(\lambda)} / j^{1/n}$, donde j es el jacobiano de la transformación.

Globalmente, el procedimiento busca un conjunto de parámetros de transformación, λ , para el cual a, b y c se satisfacen simultáneamente -al menos aproximadamente- y la información muestral sobre los tres aspectos entra en la selección de la transformación. El método depende pues de los supuestos específicos, pero sería un grave error considerar éstos hasta el final si se desea una buena aplicación.

Supongamos que tenemos un vector de n observaciones de la variable dependiente $y = (y_1, y_2, \dots, y_n)$. Los autores trabajan con una familia paramétrica de transformaciones de y a $y^{(\lambda)}$, donde el parámetro λ -que puede ser un vector- define una transformación particular y consideran dos familias de transformaciones que incluyen a las transformaciones más usuales (como la raíz cuadrada, la recíproca, la logarítmica, etc.):

$$y^{(\lambda)} = \begin{cases} (y^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \lim_{\lambda \rightarrow 0} \frac{y^\lambda - 1}{\lambda} = \ln y, & \lambda = 0 \end{cases} \quad \dots(1)$$

$$y^{(\lambda)} = \begin{cases} \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1}, & \lambda_1 \neq 0 \\ \lim_{\lambda \rightarrow 0} \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} = \ln(y + \lambda_2), & \lambda_1 = 0 \end{cases} \quad \dots(2)$$

(1) es válida para $y > 0$ y (2) para $y > -\lambda_2$. Como el análisis de la varianza no cambia con una transformación lineal, (1) es equivalente a:

$$y^{(\lambda)} = \begin{cases} y^\lambda, & \lambda \neq 0 \\ \ln y, & \lambda = 0 \end{cases} \quad \dots(3)$$

El método supone que se puede ajustar el modelo lineal:

$$E(\underline{y}^{(\lambda)}) = \underline{x} \underline{\beta} \quad \dots(4)$$

donde $\underline{y}^{(\lambda)}$ es el vector, de $n \times 1$, de las observaciones transformadas; \underline{x} una matriz conocida de dimensión $n \times p$ y $\underline{\beta}$ un vector, de $p \times 1$, de parámetros desconocidos asociado con las observaciones transformadas. Se supone también, como ya se dijo, que para algún valor desconocido de λ se cumple que las observaciones transformadas son independientes, se distribuyen Normalmente con varianza constante σ^2 y esperanza dada por (4).

Entonces la función de verosimilitud con respecto a las observaciones originales es:

$$L = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} (\mathbf{Y}^{(\lambda)} - \mathbf{x}\hat{\beta})' (\mathbf{Y}^{(\lambda)} - \mathbf{x}\hat{\beta})\right\} \cdot J \quad \dots (5)$$

donde $J = \prod_{i=1}^n \left| \frac{d y_i^{(\lambda)}}{d y_i} \right|$ es el jacobiano de la transformación.

Para determinar el valor adecuado de λ podemos proceder de dos formas: aplicando máxima verosimilitud o utilizando el método bayesiano.

a) Obtención de λ mediante máxima verosimilitud.

Para una λ fija y si \mathbf{x} es de rango completo -rango p -, por el método de máxima verosimilitud obtenemos:

$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{Y} \quad \text{y} \quad \hat{\sigma}^2(\lambda) = \frac{1}{n} (\mathbf{Y}^{(\lambda)} - \mathbf{x}\hat{\beta})' (\mathbf{Y}^{(\lambda)} - \mathbf{x}\hat{\beta})$$

donde $\hat{\sigma}^2(\lambda)$ es el estimador de σ^2 para la λ fija.

$$\text{Entonces:} \quad \text{Max} \left\{ \ln L(\lambda) \right\} = -\frac{n}{2} \ln \hat{\sigma}^2(\lambda) + \ln J + K,$$

$$\text{donde } K = -\frac{n}{2} \ln 2\pi - \frac{n}{2}$$

Para el caso particular de la transformación (1):

$$\ln J = (\lambda - 1) \sum_{i=1}^n \ln y_i$$

y de la transformación (2):

$$\ln J = (\lambda_1 - 1) \sum_{i=1}^n \ln (y_i + \lambda_2)$$

Un método para escoger la λ adecuada es calcular el valor de

$$P(\lambda) = -\frac{n}{2} \ln \hat{\sigma}^2(\lambda) + \ln J. \text{ para cada } \lambda, \text{ y obtener una gráfica}$$

de λ contra $P(\lambda)$. De esta gráfica obtenemos el valor de λ tal que $P(\lambda)$ sea máxima (es decir, tal que $\ln L(\lambda)$ es máximo).

Otro método más preciso es determinar numéricamente el valor $\hat{\lambda}$ para el cual todas las derivadas con respecto a λ son cero. En el caso particular de la transformación $y^{(\lambda)} = (y^\lambda - 1) / \lambda$, tenemos (ver Guerrero y Hernández, 1974, pp. 71-72)

$$\frac{\partial}{\partial \lambda} \text{Max} \left\{ \ln L(\lambda) \right\} = -n \frac{Y^{(\lambda)' A U^{(\lambda)}}}{Y^{(\lambda)' A Y^{(\lambda)}}} + \frac{n}{\lambda} + \sum_{i=1}^n \ln y_i$$

donde $A = I - X(X'X)^{-1}X'$, y $U^{(\lambda)}$ es el vector con componentes

$$u_i^{(\lambda)} = \lambda^{-1} y_i^\lambda \ln y_i.$$

Igualando a cero esta expresión se obtiene una ecuación cuya solución numérica² nos da la λ adecuada.

Estos resultados pueden expresarse muy simplemente si se trabaja con la transformación normalizada:

$$z^{(\lambda)} = Y^{(\lambda)} / J^{1/n}$$

Entonces tendríamos:

$$\text{Max} \left\{ \ln L(\lambda) \right\} = -\frac{n}{2} \ln \hat{\sigma}^2(\lambda; z) + \text{constante}$$

donde $\hat{\sigma}^2(\lambda; z) = \frac{z^{(\lambda)' A z^{(\lambda)}}}{n} = \frac{S(\lambda, z)}{n}$, es decir

$S(\lambda, z)$ es la suma de cuadrados de los residuales de $z^{(\lambda)}$ y, por tanto, la λ adecuada es aquella que minimiza la expresión $S(\lambda, z)$.

2. Ver p. ej. Salvadori y Baron, Numerical Methods in Engineering, Prentice-Hall, 1959.

Para la transformación (3): $z^{(\lambda)} = \frac{y^\lambda - 1}{\lambda \bar{y}^{\lambda-1}}$,

donde $\bar{y} = \left(\prod_{i=1}^n y_i \right)^{\frac{1}{n}}$ es la media geométrica de las observaciones.

Para la transformación (2): $z^{(\lambda)} = \frac{(y+\lambda_2)^{\lambda_1-1}}{\lambda_1 \{mg(y+\lambda_2)\}^{\lambda_1-1}}$

donde $mg(y+\lambda_2)$ es la media geométrica^{2a} de las observaciones trasladadas por λ_2 .

b) Obtención de λ mediante análisis Bayesiano.

El procedimiento consiste en encontrar la función de densidad de β, σ^2 y λ dada $\underline{y}^{(\lambda)}$, es decir encontrar $P(\beta, \sigma^2, \lambda / \underline{y}^{(\lambda)})$, utilizando el teorema de Bayes a partir de la verosimilitud

$P(\underline{y}^{(\lambda)} / \beta, \sigma^2, \lambda)$, que se expresa en (5). Una vez encontrada

$P(\beta, \sigma^2, \lambda / \underline{y}^{(\lambda)})$ se puede encontrar la función de densidad marginal de λ .

Es posible encontrar una densidad a priori de λ , $P_0(\lambda)$, ya sea a través de la información a priori -cuando se cuenta con ella- o considerando la uniforme sobre el rango para el cual la función de verosimilitud es apreciable (es decir, significativamente distinta de cero). El método supone también que, en ese mismo rango, las β 's y $\log \sigma$ pueden considerarse esencialmente uniforme (ver Box y Cox, 1964, pp. 217-8)

De esta forma y sabiendo que:

$$S^2(\lambda) = \frac{\underline{y}^{(\lambda)' A \underline{y}^{(\lambda)}}}{v}, \quad \text{donde } A = I - \underline{x}(\underline{x}'\underline{x})^{-1}\underline{x}'$$

$$v = n - \text{rango}(\underline{x})$$

2-a
$$\text{Es decir } mg(y+\lambda_2) = \left(\prod_{i=1}^n (y_i + \lambda_2) \right)^{\frac{1}{n}}$$

obtenemos ³ la función de densidad marginal de λ , $P(\lambda/y)$:

$$P(\lambda/y) \approx K \frac{J^{v/n}}{S^2(\lambda)^{v/2}} P_0(\lambda)$$

donde J es el jacobiano de la transformación y K es una constante independiente de λ .

De nuevo, si trabajamos con la transformación normalizada $z^{(\lambda)} = y^{(\lambda)}/J^{1/n}$, el resultado se puede expresar de una manera más simple:

$$S(\lambda, \underline{z}) = \underline{z}^{(\lambda)} \underline{A} \underline{z}^{(\lambda)}$$

y

$$P(\lambda/y) \approx S(\lambda, \underline{z})^{-v/2}$$

de forma que la λ adecuada, una vez más, es aquella que minimiza $S(\lambda, \underline{z})$.

En la práctica, podemos graficar $S(\lambda, \underline{z})^{-v/2}$ contra λ y combinar esto con cualquier información a priori sobre λ . Cuando la densidad a priori de λ puede considerarse localmente uniforme, la distribución posterior se obtiene directamente graficando $P_U(\lambda) = K \{ S(\lambda, \underline{z}) \}^{-v/2}$, donde K es tal que el área total bajo la curva es uno.

Podemos hacer una comparación entre ambos desarrollos. El $\text{Max } \ln L(\lambda)$ y el logaritmo de la distribución de λ pueden escribirse respectivamente como:

$$\underline{L}_n(\lambda) = -\frac{1}{2} n \ln \left\{ S(\lambda, \underline{z})/n \right\}, \underline{L}_b(\lambda) = -\frac{1}{2} v \ln \left\{ S(\lambda, \underline{z})/v \right\}$$

3 Sabemos que siempre es posible conseguir que \underline{x} sea de rango completo (P) y entonces tendríamos $v=n-p$

que difieren únicamente en la sustitución de n por v . Ambos son funciones monótonas de $S(\lambda, z)$ y sus máximos ocurren al minimizar la suma de cuadrados $S(\lambda, z)$. Para una descripción general, L_m y L_b son sustancialmente equivalentes. Sin embargo, puede suceder fácilmente que el cociente v/n sea apreciablemente menor que uno, aún cuando n sea muy grande. Por tanto, en las aplicaciones la diferencia no siempre puede despreciarse.

Ejemplo.

Damos un ejemplo publicado en el artículo original en que se expone el método. Consiste en un experimento sobre el comportamiento de hilados sujetos a ciclos de cargas sucesivas para el engrosamiento de la fibra. En la tabla 1 damos el número de ciclos para obtener una falla en el proceso, según los valores de los factores siguientes:

- x_1 : longitud del espécimen probado (250, 300, 350 m.m.)
- x_2 : amplitud del ciclo de carga (8, 9, 10 m.m.)
- x_3 : carga (40, 45, 50 m.m.)

Los correspondientes valores de éstas x para cada valor de y se denotan convencionalmente por -1 , 0 y 1 .

Antes de aplicar propiamente el método, veremos que es posible deducir algunos resultados analizando la tabla.

El rango de variación de la y es sumamente amplio -de 90 a 3636 ciclos- por lo cual resultaría razonable tratar de analizar $\log y$, lo cual implicaría una gran simplificación del problema en este sentido. Box y Cox afirman que también parece natural tomar logaritmos en las variables independientes, esto es, consideran modelos del siguiente tipo:

$$y = x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot x_3^{\beta_3} \dots (1)$$

Si linealizamos este modelo tomando logaritmos en ambos miembros, es decir, si consideramos el modelo

$$\log Y = \beta_1 \log x_1 + \beta_2 \log x_2 + \beta_3 \log x_3$$

obtenemos los siguientes estimadores para las β 's, con las estimaciones de las desviaciones estándar correspondientes:

$$\hat{\beta}_1 = 4.96 \pm 0.20, \quad \hat{\beta}_2 = -5.27 \pm 0.30, \quad \hat{\beta}_3 = -3.15 \pm 0.30$$

De aquí vemos que $\hat{\beta}_1 = -\hat{\beta}_2$, lo cual sugiere que la combinación $\log x_1 - \log x_2 = \log(x_1/x_2)$ puede ser relevante.

De hecho, x_2/x_1 representa la amplitud del ciclo de carga por fracción de espécimen. Si hacemos $x_4 = x_2/x_1$ y redondeamos los coeficientes de regresión obtenidos, resulta el modelo:

$$y \approx x_4^{-5} x_3^{-3}$$

que se ajusta muy bien a los datos. De esta manera, podemos decir que hay argumentos generales -basados en el análisis simple de los datos- bastante fuertes a favor de una transformación logarítmica de todas las variables.

Será interesante ahora, ver si el método propuesto nos indica la misma transformación logarítmica. Consideramos únicamente transformaciones de la variable dependiente, ya que, además de que el método no permite otra cosa, la transformación de las x no tendría gran efecto en la linealidad puesto que el rango de sus valores es bastante pequeño.

Consideramos la familia de transformaciones (3), es decir:

$$y^{(\lambda)} = \begin{cases} y^\lambda, & \lambda \neq 0 \\ \ln y, & \lambda = 0 \end{cases}$$

y, en términos de la variable estandarizada $z^{(\lambda)}$:

$$z^{(\lambda)} = \begin{cases} \frac{y^{\lambda-1}}{\lambda \hat{y}^{\lambda-1}}, & \lambda \neq 0 \\ \lim_{\lambda \rightarrow 0} \frac{y^\lambda - 1}{\lambda \hat{y}^{\lambda-1}} = \hat{y} \ln y, & \lambda = 0 \end{cases}$$

donde, como ya vimos, $\hat{y} = \left(\prod_{i=1}^n y_i \right)^{1/n}$.

El método supone que después de esta transformación:

- a) El valor esperado de la variable dependiente -i.e., el μ_0 delo- puede ser representado por un modelo lineal en las x ;
- b) La varianza de los errores es constante; y
- c) Las observaciones se distribuyen Normalmente.

Estos supuestos pueden ser verificados mediante las técnicas usuales (prueba de Shapiro y Wilk, prueba de Durbin-Watson, prueba de Levene, análisis de residuales, etc.) y, además, en el capítulo 4 describimos algunos procedimientos para verificar la utilidad de la transformación.

La función de verosimilitud maximizada y la distribución marginal de λ están en función de la suma de cuadrados de los residuales de $z^{(\lambda)}$, $S(\lambda, \underline{z})$. Puesto que hay 4 constantes en el modelo de regresión lineal y 27 observaciones, la suma de cuadrados tiene $27-4=23$ grados de libertad. Recordando que $S(\lambda, \underline{z}) = \sum_1 (z_1^{(\lambda)} - \hat{z}_1^{(\lambda)})^2$, la tabla 2 muestra los valores de $S(\lambda, \underline{z})$, $\text{Max. ln}\{L(\lambda)\}$ y $P_u(\lambda)$ para distintos valores de λ y graficamos los resultados en la gráfica 1. Los cálculos fueron hechos en base a las fórmulas siguientes:

$$L_m(\lambda) = -\frac{n}{2} \ln \hat{\sigma}^2(\lambda, \underline{z}) = -13.5 \ln \hat{\sigma}^2(\lambda, \underline{z}) = \ln \{S(\lambda, \underline{z})\}^{-13.5} \quad +44.49$$

Para $P_u(\lambda)$ hacemos uso de $L_b(\lambda) = -\frac{v}{2} \ln \{S(\lambda, \underline{z})/v\}$, de manera que

$$P_u(\lambda) = K \exp\{L_b(\lambda)\}$$

donde, como vimos, K es el recíproco del área bajo la curva ⁴

$y = \exp. L_b(\lambda)$, entonces en este caso:

$$P_u(\lambda) = K \exp \left(\frac{\ln S(\lambda, z)^{-v/2}}{v} \right) = 0.540 \times 10^{-6} \cdot S(\lambda, z)^{-11.5}$$

T A B L A 2

λ	$S(\lambda, z)$	$L_m(\lambda)$	λ	$S(\lambda, z)$	$L_m(\lambda)$
1.00	5.4810	21.52	-0.20	0.2920	61.11
0.80	2.9978	29.67	-0.40	0.5478	52.61
0.60	1.5968	38.17	-0.60	1.1035	43.16
0.40	0.8178	47.21	-0.80	2.1396	34.22
0.20	0.4115	56.48	-1.00	3.9955	25.79
0.00	0.2519	63.10			

λ	$P_u(\lambda)$	λ	$P_u(\lambda)$
0.20	0.02	-0.10	4.66
0.15	0.09	-0.15	2.36
0.10	0.42	-0.20	0.77
0.05	1.58	-0.25	0.19
0.00	4.18	-0.30	0.04
-0.05	5.64	-0.35	0.01

⁴ El área bajo la curva $y = \exp L_b(\lambda)$ fue determinada por integración numérica. (ver p. ej. Salvadori y Baron, op. cit., p.70)

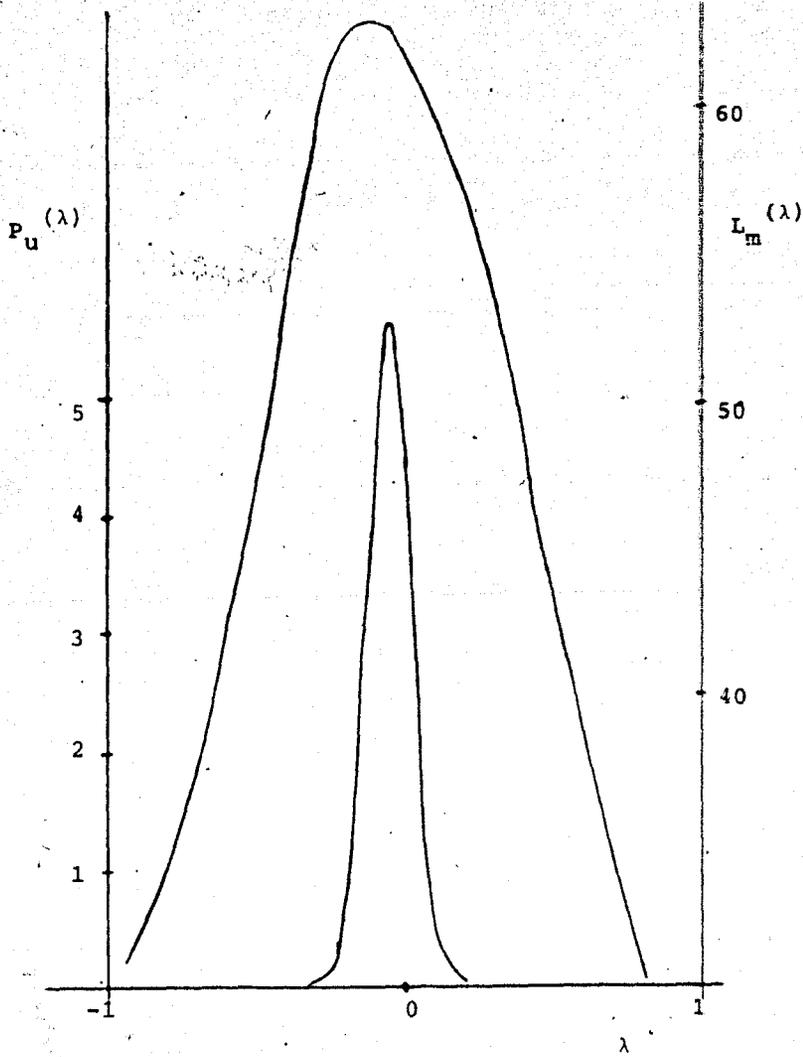


FIGURA 1

Analizando con más detalle el intervalo de 0.00 a -0.20 para el caso de $L_m(\lambda)$, y el intervalo (-0.05, -0.10) para el caso de $P_u(\lambda)$ los autores concluyen que el valor óptimo de λ es aproximadamente $\hat{\lambda} = -0.06$. El intervalo de confianza del 95% va, aproximadamente, de -0.18 a 0.06. La distribución de $P_u(\lambda)$ tiene su media en -0.06 y alrededor del 95% de la distribución está entre -0.20 y 0.08.

La conclusión es, pues, que la selección de $\lambda=0$ -que implica una transformación logarítmica- tiene grandes ventajas y está fuertemente fundamentada por los datos.

2.2. TRANSFORMACION DE LA VARIABLE INDEPENDIENTE METODO DE BOX Y TIDWELL (1962)

Los autores suponen que los errores se distribuyen Normalmente al menos de una manera aproximada, independientemente y con varianza constante y se concentran en encontrar transformaciones en la variable independiente, x , para obtener una función de regresión tan simple como sea posible. Sin embargo, la aplicación del método no se limita a este aspecto. Quizá su uso más general es cuando no se conoce la forma de la función que relaciona a las variables indepen-

dientes con la variable dependiente. En estos casos, cuando se desconoce la verdadera relación funcional $f(\underline{x}, \theta)$, con frecuencia puede progresarse considerando una función $g(\underline{x}, \beta)$ -a la que se llama función "graduante"-, la cual puede esperarse que represente bien a la verdadera relación en cierta región del espacio \underline{x} . β está compuesta de p constantes empíricas, donde $1 \leq p$ - p es el número de parámetros de θ -. En particular, a menudo se intenta que $g(\underline{x}, \beta)$ sea un polinomio de 1º o 2º grado en \underline{x} . Esta técnica para el estudio empírico de las características locales de la función "de respuesta" ha sido llamada "Análisis de la superficie de respuesta" y se ha aplicado ampliamente en el trabajo experimental, particularmente en la industria química.

El propósito fundamental es, entonces, trabajar con un polinomio de grado "pequeño" en las variables transformadas en vez de un polinomio de mayor grado en las variables originales.

Las ventajas de aproximar la expresión mediante una forma lineal -que es factible si la función es estrictamente monótona- son las siguientes:

- a) Se pueden apreciar más fácilmente sus implicaciones; y
- b) Simplifica los cálculos subsecuentes para los que esta expresión pudiera necesitarse.

Un uso común del Análisis de superficie de respuestas es en-

contrar un conjunto de valores $x_{10}, x_{20}, \dots, x_{k0}$ de las variables independientes, tales que una serie de j variables, dependientes de x_1, x_2, \dots, x_k , estén dentro de ciertos límites especificados. Cuando pueden encontrarse unas funciones lineales aproximadas -grduantes- en las variables transformadas adecuadamente, el problema de satisfacer simultáneamente esta j desigualdades, es un problema estándar de programación lineal y puede resolverse con las técnicas ordinarias. El correspondiente problema cuando las funciones no sean lineales es más difícil.

Por otra parte, cuando se analiza a la función $f(\underline{x}, 0)$ en una región en que tiene un extremo, en ocasiones podremos representarla mediante un polinomio de 2º grado, problema relativamente fácil (Box y Wilson, 1951). Si el polinomio de 2º grado en las variables originales no es adecuado, es preferible hacer una transformación para obtener un adecuado polinomio de 2º grado en las variables transformadas que ajustan un polinomio de mayor grado sin transformar.

El procedimiento considera a cada variable transformada en función de sólo una variable original, aunque lo ideal sería que estuviera en función de todas las variables originales.

Supongamos que contamos con las observaciones y_1, y_2, \dots, y_n y los valores de las n variables x_1, x_2, \dots, x_n , donde x_i es un vector de $k \times 1$ que indica los valores de las x para la i -ésima observau

ción. Suponga además que:

$$E(Y_i) = \eta_i \quad \text{y} \quad E(Y_i - \eta_i)(Y_j - \eta_j) = \begin{cases} \sigma^2, & \text{si } i=j \\ 0, & \text{si } i \neq j \end{cases}$$

donde σ^2 es desconocida. Supongamos que η puede representarse aproximadamente en la región de interés mediante una función simple

$\eta = f(\underline{h}, \underline{g})$, donde los elementos h_1, h_2, \dots, h_k del vector \underline{h} son las x transformadas de alguna forma adecuada de modo que

$h_i = h_i(\underline{x}_i, \underline{a}_i)$, donde \underline{x}_i es un vector de dimensión P_i , cuyos elementos son las constantes desconocidas de la transformación. \underline{g} es

un vector de l constantes desconocidas que dependen de las transformaciones particulares -i.e., dependen de la selección de las

α 's-. Supongamos que $\alpha_i^{(0)} = (\alpha_{i1}^{(0)}, \alpha_{i2}^{(0)}, \dots, \alpha_{ip_i}^{(0)})$ son los valores iniciales de las constantes de la i -ésima observación. En la

1a. iteración, estos valores podrían ser tales que $h_i = x_i$, es decir, tales que no se ha aplicado ninguna transformación. Podemos

determinar $h_u^{(0)}$ para el vector cuyo i -ésimo elemento es

$h_i(x_{iu}, \alpha_i^{(0)})$. Desarrollando en estos valores supuestos e ignorando los términos de grado mayor que uno en $(\alpha_{ij} - \alpha_{ij}^{(0)})$ tenemos

aproximadamente:

$$\eta_u = f(h_u^{(0)}, \underline{g}) + \sum_{i=1}^k \sum_{j=1}^{P_i} (\alpha_{ij} - \alpha_{ij}^{(0)}) \left\{ \frac{\partial f(\underline{h}, \underline{g})}{\partial \alpha_{ij}} \right\}_{\substack{h=h_u^{(0)} \\ \alpha_i = \alpha_i^{(0)}}} \dots (1)$$

$$\text{donde } \left\{ \frac{\partial f(\underline{h}, \underline{g})}{\partial \alpha_{ij}} \right\}_{\substack{h=h_u^{(0)} \\ \alpha_i = \alpha_i^{(0)}}} = \left\{ \frac{\partial f(\underline{h}, \underline{g})}{\partial h_i} \right\}_{h=h_u^{(0)}} \left\{ \frac{\partial h_{iu}}{\partial \alpha_{ij}} \right\}_{\alpha_i = \alpha_i^{(0)}}$$

Podemos calcular las cantidades $\left\{ \frac{\partial h_{iu}}{\partial \alpha_{ij}} \right\}_{\alpha_i = \alpha_i^{(0)}}$ pues conocemos la forma particular de las transformaciones h_i que hemos decidido usar, pero deberemos de estimar las cantidades

$\left\{ \frac{\partial f(\underline{h}, \underline{\beta})}{\partial h_i} \right\}_{h=h_u^{(0)}}$ de alguna manera. Esto lo podemos hacer convenientemente con un ajuste preliminar de las observaciones a $n_u = f(h_u^{(0)}, \underline{\beta})$ mediante el método de mínimos cuadrados o cualquier otro. Podemos derivar la expresión ajustada $y_u = f(h_u^{(0)}, b)$, donde b es el estimador de mínimos cuadrados de $\underline{\beta}$ para obtener valores aproximados de las cantidades requeridas $\left\{ \frac{\partial f(\underline{h}, \underline{\beta})}{\partial h_i} \right\}_{h=h_u^{(0)}}$,

Los $n \sum_{i=1}^k P_i$ valores aproximados de

$\left\{ \frac{\partial f(\underline{h}, \underline{\beta})}{\partial \alpha_{ij}} \right\}_{\substack{h=h_u^{(0)} \\ \alpha_i = \alpha_i^{(0)}}}$ que podemos ahora calcular, nos

proveen de un conjunto de variables independientes del cual podemos obtener los ajustes a las constantes de las transformaciones reajustando las observaciones a la expresión completa del miembro derecho de la ecuación (1) mediante el método de mínimos cuadrados. Estas constantes ajustadas pueden tomar ahora el lugar de los valores iniciales supuestos en los cálculos anteriores y repetir el ciclo completo.

Uno de los tipos de transformación más simple que se puede utilizar, pero que incluye a muchas de las transformaciones más usuales, es la transformación potencia:

$$h_i = \begin{cases} x_i^{\alpha_i}, & \alpha_i \neq 0 \\ \ln x_i, & \alpha_i = 0 \end{cases} \dots (2)$$

En la práctica, si hacemos suposiciones distintas de 1 para los valores de α_i en la 1a. iteración, siempre podemos considerar estas $x_i^{\alpha_i}$ como las variables básicas x_i , de tal forma que sólo consideramos el caso en que $\alpha_i = \alpha_j = 1$. Habiendo obtenido los nuevos valores del conjunto de valores iniciales, podemos tratar las resultantes x transformadas como nuevas x para llevar a cabo una segunda iteración y así sucesivamente. Tenemos entonces el modelo

$\eta = f(\underline{h}, \underline{\beta})$ con $h_i = x_i^{\alpha_i}$ y $\alpha_i^{(0)} = 1, i=1, 2, \dots, k$ para nuestros primeros supuestos.

Entonces aproximadamente:

$$\eta_u = f(\underline{x}_u, \underline{\beta}) + \sum_{i=1}^k (\alpha_i - 1) \left\{ \frac{\partial f(\underline{h}, \underline{\beta})}{\partial \alpha_i} \right\}_{\substack{h=x_u \\ \alpha_i=1}} \dots (3)$$

es decir,

$$\eta_u = f(\underline{x}_u, \underline{\beta}) + \sum_{i=1}^k (\alpha_i - 1) \left\{ \frac{\partial f(\underline{x}_u, \underline{\beta})}{\partial x_{iu}} \right\} x_{iu} \ln x_{iu} \dots (4)$$

Procedemos 1º ajustando $\hat{y}_u = f(x_u, b)$. Entonces derivamos esta expresión para obtener las cantidades $\partial f(x_u, b) / \partial x_{iu}$ que estiman las cantidades entre las llaves en (4). Multiplicamos por $x_{iu} \ln x_{iu}$ para proveer las variables independientes nuevamente construidas:

$$z_{iu} = \left\{ \frac{\partial f(x_u, b)}{\partial x_{iu}} \right\} x_{iu} \ln x_{iu}$$

Reajustando por mínimos cuadrados las cantidades:

$$n_u = f(x_u, \underline{\beta}) + \sum_{i=1}^k (\alpha_i - 1) z_{iu}$$

obtenemos los estimadores $\hat{\alpha}_i - 1$ de $\alpha_i - 1$. Podemos considerar ahora a las variables $x_i^{\alpha_i}$ como las x_i en una iteración posterior.

Si deseamos ajustar una función lineal en las h_i , es decir, si:

$$n_u = f(h_u, \underline{\beta}) = \beta_0 + \sum_{i=1}^k \beta_i h_{iu}$$

donde $h_{iu} = x_{iu}^{\alpha_i}$, entonces, siguiendo el procedimiento anterior primero ajustamos:

$$\hat{Y}_u = b_0 + \sum_{i=1}^k b_i x_{iu}$$

Podemos ahora reajustar:

$$\hat{Y}_u' = b_0' + \sum_{i=1}^k b_i' x_{iu} + \sum_{i=1}^k c_i x_{iu} \ln x_{iu}$$

y los primeros estimadores de las α 's están dados por $\hat{\alpha}_i = (c_i/b_i) + 1$.

El proceso puede entonces repetirse con las nuevas variables $x_i^1 = x_i^{\hat{\alpha}_i}$. Esta aplicación particular ha sido investigada por Turner, Monroe y Lucas (1961).

Una característica importante del procedimiento es su rapidez de convergencia, aunque quizá una desventaja es que debe ponerse especial atención en los errores de redondeo, pues se ha visto en algunos ejemplos que el no considerar un número suficiente de decimales acarrea violentas oscilaciones en las iteraciones sucesivas.

Es factible esperar estimadores razonablemente buenos de la transformación requerida cuando $q = 0.5 r/m$ no es muy pequeña o σ no es muy grande. El procedimiento debe detenerse al obtener un ajuste suficientemente bueno.

Ejemplo:

Damos un ejemplo publicado en el artículo original en el que $k = 1$, es decir, se tiene sólo una variable independiente. Siguiendo el procedimiento visto y si utilizamos la transformación potencia tendríamos:

$$n_u = f(h_u, g) = \beta_0 + \beta_1 h_u = \beta_0 + \beta_1 x_u^\alpha$$

Tomando como valor inicial $\alpha = 1$, primero ajustamos el modelo:

$$\hat{y}_u = b_0 + b_1 x_u$$

y luego

$$\hat{y}_u' = b_0' + b_1' x_u + c_1 x_u \ln x_u$$

y los primeros estimadores de α están dados por $a_1 = \frac{c_1}{b_1} + 1$.

Podemos repetir el proceso con las nuevas variables $x^i = x^{a_1}$.

⁵ Se supone que se tienen n valores de la variable independiente original, x , que se encuentran igualmente espaciados, con media m y rango r .

El ejemplo numérico es el siguiente:

Un investigador observa la cantidad de polímeros ⁶ formados en una hora bajo condiciones estables en varias cargas igualmente espaciadas de un catalizador. Los datos son los siguientes:

T A B L A 3		
Gramos de catalizador cargados al reactor	Gramos del polímero formado	Valores de W para $q=0.6$ y $n=5$
x	y	
0.5	28.45	0.321 041
1.5	47.78	-0.207 594
2.5	64.35	-0.281 630
3.5	74.67	-0.098 120
4.5	83.65	0.266 304

$$\text{Divisor } \Delta = 0.195854$$

Posteriormente explicaremos el significado de la 3a. columna y del divisor Δ .

De acuerdo con estos datos el primer ajuste resulta

$$\hat{y} = 25.4575 + 13.729 x$$

con un coeficiente de correlación del 98.62%.

Luego tendríamos que reajustar $\hat{y} = b_0' + b_1'x + c_1x \ln x$

En la práctica, en algunas ocasiones es conveniente calcular c utilizando las funciones ortogonales que podemos obtener de la siguiente manera:

⁶ Un polímero puede definirse como una cadena de moléculas o como una molécula "grande" que resulta de la reacción de muchas moléculas sencillas (monómeros) capaces de combinarse en series.

Supongamos que se cuenta con n valores de la variable independiente original x , igualmente espaciados, con media m y rango r . Podemos definir entonces la cantidad:

$$q = \frac{1}{2} \frac{\text{rango de la variable independiente}}{\text{media de la variable independiente}} = \frac{1}{2} \frac{r}{m}$$

y, por medio de ésta, la cantidad $W = [x \ln x] / m q^2$, donde $[x \ln x]$ representa la parte ortogonal de $x \ln x$. En el apéndice se dan los valores de W para distintos valores de q y de n . En nuestro ejemplo $m=2.5$ y $r=4$, entonces $q=0.8$. En las tablas de las funciones ortogonal encontramos los valores de W para $q=0.8$ y $n=5$ y el correspondiente valor del divisor Δ que reproducimos en la tabla 3. Podemos entonces calcular c con la fórmula:

$$c = \sum_i W_i y_i / m \Delta$$

y, por tanto, en nuestro ejemplo obtenemos $c = -8.0844$.

De donde a , el estimador de a , es:

$$a = \frac{c}{b} + 1 = \frac{-8.0844}{13.729} + 1 = 0.4111$$

Podemos hacer ahora nuevas iteraciones tomando como variable independiente $x_u^a = x_u^a$. En este caso obtenemos:

Iteración	a	b
2a.	0.458	44.306
3a.	0.474	42.443

y entonces el modelo final es:

$$\hat{y} = -2.411 + 42.443 x^{0.474}$$

que tiene una correlación casi perfecta del 99.91% .

Podemos considerar que los datos sugieren que una transformación conveniente podría ser $x^{1/2}$, ya que una alteración tan pequeña del valor de α estimado puede ser despreciable.

METODO DE DOLBY (1963).

El objetivo primordial de este método es conseguir que la variable dependiente esté relacionada linealmente con la variable independiente (teniendo en cuenta nuevamente, que una transformación que linealiza la función puede servir también para aproximarnos más a los otros supuestos) y encontrar la relación lineal que mejor se ajuste a los datos.

El autor se limita a considerar la familia de transformaciones de la forma $T(x) = (c + x)^p$, donde p es un número real y c es lo suficientemente grande para lograr que $c+x > 0$ en el recorrido de los datos. Se trataría entonces de ajustar el modelo

$$y = a + b (c + x)^p \quad \dots (1)$$

El procedimiento consiste en examinar una ecuación diferencial que caracteriza a la familia de transformaciones y de ahí deriva una ecuación -por el método de diferencias finitas- que es útil

para obtener un estimador de p , que podemos usar como una solución aproximada o como el valor inicial de una solución más precisa.

Si y está dada por (1), podemos obtener la ecuación diferencial:

$$\frac{d}{dx} \left(\frac{y'}{y''} \right) = \frac{1}{p-1} \quad \dots (2)$$

Sea f una función que es una solución de esta ecuación diferencial, entonces podemos definir:

$$A_f = 0.5 + \frac{d}{dx} \left(\frac{f'}{f''} \right) \quad \dots (3)$$

que servirá como un índice de la familia de transformaciones. Algunas propiedades de A_f sugieren una relación de correspondencia de la recta real con el círculo, con $A_f = 0$ en el polo sur de éste y $A_f = \pm \infty$ en el polo norte. Bajo tal correspondencia y si escogemos el diámetro del círculo igual a $\sqrt{3}/2$, las transformaciones de mayor interés práctico estarán igualmente espaciados alrededor del círculo. En la figura 2 mostramos la gráfica resultante.

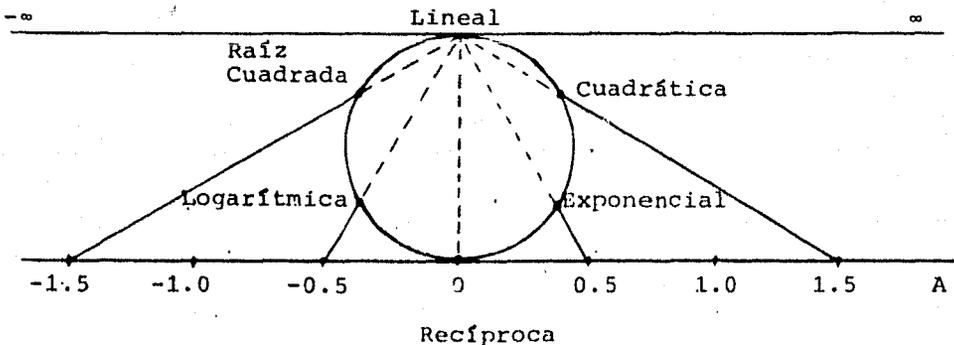


FIGURA 2

Por el método de diferencias finitas es posible calcular una aproximación a la ecuación diferencial (3), de forma que:

$$A_f = 0.5 + \frac{d}{dx} \left(\frac{f^1}{f^{11}} \right) = 0.5 + \left\{ \frac{0.5(d_3 + d_2)}{d_3 - d_2} - \frac{0.5(d_2 + d_1)}{d_2 - d_1} \right\} \equiv a_f$$

donde $d_i = f(x_{i+1}) - f(x_i)$ y suponemos que los valores de la variable independiente están igualmente espaciados.

Tenemos, pues, una aproximación, a_f , al valor A_f , obtenida de los datos. Podemos entonces proyectar el valor de a_f del eje real al polo norte del círculo para determinar cuál de las 6 transformaciones consideradas debemos usar.

Si los valores de la variable independiente no se encuentran igualmente espaciados, hemos de definir a_f de otra forma. Entonces, si

$$[i+1, i] = \frac{d_i}{x_{i+1} - x_i}$$

$$y, \quad [i+2, i+1, i] = \frac{d_{i+1}}{(x_{i+2} - x_i)(x_{i+2} - x_{i+1})} - \frac{d_i}{(x_{i+2} - x_i)(x_{i+1} - x_i)}$$

podemos definir:

$$A_f = \frac{1}{2} \left\{ 1 + \frac{1}{x_4 + x_3 - x_2 - x_1} \left(\frac{[4, 3] + [3, 2]}{[4, 3, 2]} - \frac{[3, 2] + [2, 1]}{[3, 2, 1]} \right) \right\}$$

Es importante hacer notar que en cualquiera de los dos casos sólo

utilizamos 4 valores de la variable independiente para determinar a_f . Estos valores deben ser los extremos de las observaciones ordenadas -es decir, el valor menor y mayor de x - y los dos puntos interiores más cercanos a aquéllos que subdividen el intervalo en 3 intervalos iguales; esto es, debemos dar a x_2 el valor de la x más aproximado a (ver Dolby, 1963, p. 323):

$$x_{(2)} = x_1 + \frac{(x_4 - x_1)}{3}$$

y, análogamente, dar a x_3 el valor de x más próximo a:

$$x_{(3)} = x_1 + \frac{2(x_4 + x_1)}{3}$$

Con esto tenemos entonces una solución aproximada para elegir el valor de p .

De las ecuaciones (2) y (3) y de la aproximación a A_f se puede obtener una solución más rigurosa, puesto que:

$$\frac{1}{p-1} = \frac{0.5(d_3+d_2)}{d_3-d_2} - \frac{0.5(d_2+d_1)}{d_2-d_1}$$

es decir, $\frac{1}{p-1} = a_f - 0.5$, obtenemos la aproximación

$$p = \frac{1}{a_f - 0.5} + 1$$

El cálculo de a_f puede hacerse también dividiendo las observaciones en subconjuntos de 4 puntos, de cada uno de los cuales se

obtendrá un estimador de a_f , que luego proveerían un estimador final como resultado del promedio de ellos. Sin embargo, de esta manera, el error de cada estimador de a_f sería incrementado al menos en n y al promediar estos n estimadores independientes -suponiendo que el conjunto de observaciones se puede subdividir en n subconjuntos, cada uno con 4 puntos-, el error sólo se reduciría en $n^{1/2}$, lo cual conduciría a una pérdida considerable. Es muy dudoso, pues, que esta forma de efectuar los cálculos valga la pena.

Dolby afirma que parece razonable concluir que cuando las desviaciones no son muy grandes en el caso de que los valores de la variable independiente no estén igualmente espaciados y si la razón del error experimental al rango de la variable dependiente es del 2% o menos, entonces el método es útil como un indicador aproximado de la transformación adecuada. Sin embargo, él basa estas afirmaciones -principalmente la segunda- en los ejemplos que ha resuelto en los cuales efectivamente el método funciona bien.

Una de las mayores ventajas del método es su rapidez, puesto que, además de que, una vez que se conocen c y p , la determinación de a y b es directa (y la determinación de c no es difícil), el procedimiento mismo para determinar el valor de p es bastante rápido. Sin embargo, el método gráfico puede no ser muy satisfactorio en algunas ocasiones, a pesar de que en los ejemplos dados en el artículo original producen buenos resultados. Además, como es claro, otra des

ventaja del procedimiento es que sólo es útil cuando se tiene una sola variable independiente.

Ejemplo:

Aplicamos este procedimiento a los mismos datos del ejemplo utilizado con el método de Box y Tidwell.

Primero hemos de determinar los cuatro valores de la variable independiente que entrarán en los cálculos. Sabemos que debemos tomar $x_1=1.5$ y $x_4=4.5$, x_2 deberá ser el valor más cercano a:

$$x_{(2)} = x_1 + \frac{(x_4 - x_1)}{3} = 0.5 + \frac{4}{3} = 1.833$$

y x_3 el más cercano a:

$$x_{(3)} = x_1 + \frac{2(x_4 - x_1)}{3} = 0.5 + \frac{8}{3} = 3.166$$

por tanto, $x_2 = 1.5$ y $x_3 = 3.5$. De esta forma tenemos:

x_i	$f(x_i)$	d_i
0.5	28.45	
1.5	47.78	19.33
3.5	74.67	26.89
4.5	83.65	8.98

de donde $a_f = -3.558$. Proyectando este valor del eje real al polo norte del círculo, obtenemos la siguiente gráfica:

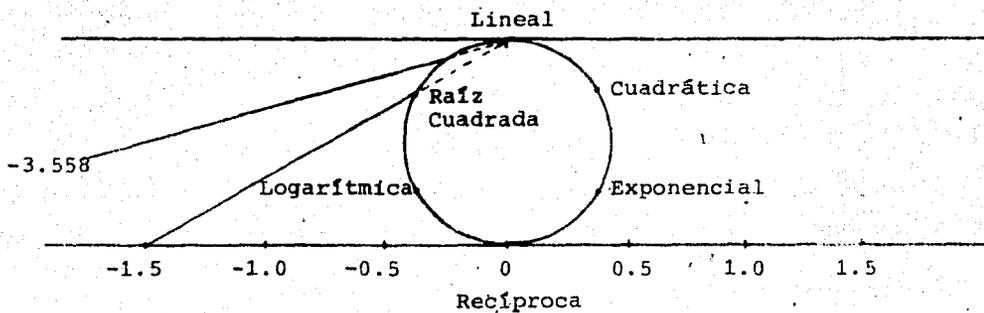


FIGURA 3

Por tanto, según la gráfica, la transformación sugerida es ta entre la \sqrt{x} y x . Con el método de Box y Tidwell, la transformación final resultaba aproximadamente la \sqrt{x} , de manera que el modelo sería semejante, aunque en realidad la proyección se inclina más a la transformación lineal (o más propiamente, a no hacer ninguna transformación) que a la raíz cuadrada.

Si deseamos un estimador más riguroso, obtendríamos:

$$p = \frac{1}{a_f - 0.5} + 1 = 0.7536$$

Podemos tomar $c=0$, pues así se cumple que $c+x > 0$ en todo el recorrido. Entonces, el modelo obtenido es:

$$y = a + b x^{0.7536}$$

que difiere del obtenido con la proyección y con el de Box y Tidwell,

aunque la diferencia no es alarmante. Podríamos concluir que ésta última estimación de p sugiere el uso de $x^{3/4}$ como variable independiente, resultando el modelo:

$$\hat{y} = 16.9998 + 22.312 x^{0.75}$$

con una correlación muy buena del 99.53%, aunque ligeramente menor que la obtenida con el método precedente.

2.3 TRANSFORMACION DE AMBAS VARIABLES

METODO DE DRAPER Y HUNTER (1969)

El procedimiento consiste en ayudar en la selección de una transformación mediante gráficas de funciones que se dan naturalmente en el análisis usual. Cuando hacemos varias de estas gráficas, puede suceder que no todas indiquen la misma transformación y esto es lo fundamental del método: analizar, de manera gráfica, las consecuencias de hacer varias transformaciones a un conjunto de datos.

El hecho de que no todas las gráficas indiquen la misma transformación, puede parecer, a primera vista, una desventaja pero la realidad no es así: usualmente no se puede obtener una sola transformación que satisfaga varios criterios distintos. Lo que el método pretende es analizar simultáneamente una serie de criterios diferentes para poder tomar una actitud más flexible. Como

los criterios que hemos de considerar dependen del problema particular, no hay ninguno que sea fijo, ni hay tampoco una metodología que aplicar en todos los casos salvo el analizar las gráficas de algunas estadísticas que estén en función de los distintos valores de los parámetros que definen la transformación considerada.

Como la eficiencia del método estriba en el análisis simultáneo de diversos aspectos o criterios, el método supone la ayuda de un buen equipo de computación que permita la elaboración de las gráficas y el estudio simultáneo de aquéllos. Normalmente se podrá hacer uso de los programas estándar existentes para regresión y análisis de la varianza, con un pequeño programa introductorio evitando la elaboración de programas especiales.

Uno de los aspectos más atractivos del procedimiento es, quizá, la flexibilidad del mismo, ya que permite no sólo el estudio de una transformación tanto en la variable dependiente como en la independiente, sino el análisis de transformaciones diversas. Incluso es posible considerar también al mismo tiempo algunos de los métodos presentados anteriormente. Trata de satisfacer asimismo los tres supuestos asociados con los modelos lineales: aditividad, varianza constante y Normalidad, con la modalidad de que las estadísticas para verificar estos supuestos pueden incluirse en el mismo procedimiento de selección, evitando así posteriores verificaciones.

Ejemplos.

Presentamos dos ejemplos con el fin de resaltar la importancia y flexibilidad de este procedimiento.

1) Reproducimos un ejemplo original de Turner, Monroe y Lucas(1959), en el que los datos consisten en tiempos récord y sus correspondientes velocidades, observados en carreras de atletismo.

T A B L A 4

Tiempo	Velocidad
x	y
(seg)	(m/seg)
20.0	10.06
45.2	8.85
105.7	7.57
139.0	7.19
237.8	6.77
302.2	6.62
472.8	6.35
816.8	6.12
1710.4	5.85
3600.0	5.57

Supongamos que deseamos transformar los datos mediante $z=x^a$ para ajustar el modelo de primer orden

$$E(y) = \beta_0 + \beta_1 z$$

Esto nos llevaría a realizar un análisis de la varianza del siguiente tipo:

Fuente	Grados de Libertad	Suma de Cuadrados	Cuadrado Medio	Cociente
b_0	1	scb_0	cmb_0	
b_1/b_0	1	scb_1	cmb_1	$m=cmb_1/s^2$
Residual	8	scr	s^2	
Total	10	sct		

Debido al modelo que queremos ajustar deberemos escoger 'a' de manera que cmb_1 sea grande y s^2 sea pequeña (es importante hacer notar que si transformáramos la variable dependiente y , deberíamos examinar los cocientes de los cuadrados medios y no los cuadrados medios, puesto que la suma de cuadrados total cambia con las transformaciones de y , lo cual no sucede en este caso). Dado que scb_0 y sct son constantes, scr se minimiza cuando scb_1 es máxima. Entonces s^2 es mínima cuando cmb_1 es máxima y, por tanto, nos bastará con examinar la gráfica de s^2 contra a . Recordando nuevamente que $s^2 = \sum_i (y_i - \hat{y}_i)^2$ y para valores de 'a' en el intervalo de -3.0 a 2.0 obtenemos:

T A B L A 5

a	s^2	a	s^2
-3.0	0.96254	-0.42	0.00868
-2.5	0.87541	-0.415	0.00864
-2.0	0.74272	-0.41	0.00867
-1.5	0.54297	-0.4	0.00892
-1.0	0.26367	0.0	0.26803
-0.5	0.01744	0.5	1.01817
-0.45	0.01022	1.0	1.55388
-0.44	0.00946	1.5	1.81600
-0.43	0.00895	2.0	1.94693

Mostramos la gráfica de los resultados en la siguiente figura:

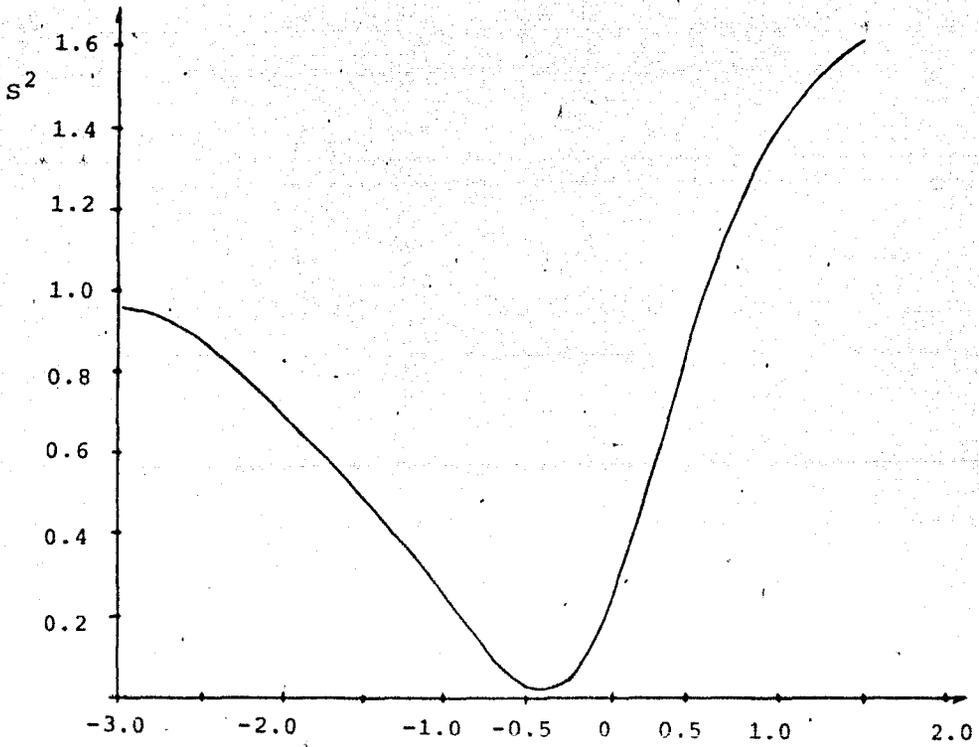


FIGURA 4

a

De aquí vemos que s^2 alcanza su mínimo aproximadamente en $a = -0.41$. Si no utilizáramos ninguna transformación, el valor de s^2 sería el correspondiente a $a=1$ y el grado de mejoría para cualquier valor particular de a , digamos a_0 , puede obtenerse al comparar el valor de s^2 en $a=a_0$ con el de s^2 en $a=1$. De la gráfica podemos afirmar que el valor de $a=-0.5$, si bien no es óptimo, es un valor bastante cercano a éste y tiene la ventaja de que se puede manejar más fácilmente. Si consideramos, pues, el valor $a=-0.5$, obtenemos el modelo

$$\hat{y} = 5.341 + \frac{21.957}{\sqrt{x}}$$

con una correlación del 99.65%.

Por último, el método recomienda el uso del análisis de residuales (ver p. ej. Draper y Smith, 1966, pp. 86-100), que nos facilita la verificación de los supuestos usuales del análisis de la varianza. En este ejemplo, los residuales, según el orden de las observaciones, son las siguientes:

-0.21, 0.24, 0.09, -0.01, 0, 0.02, 0, 0.01, -0.02, -0.14

de donde podemos ver que el ajuste es menos bueno en los extremos; sin embargo, el residual más grande es sólo el 2.7% de la observación correspondiente; por tanto, podemos afirmar que el modelo nos provee de una representación muy satisfactoria de los datos.

2) El siguiente es un ejemplo original de Box y Cox (1964) que consiste en los tiempos de supervivencia de animales (dados en unidades de 10 hs.) en un experimento factorial de 3×4 (3 venenos y 4 tratamientos). Cada combinación de los dos factores -12 en total- se aplica a 4 animales. La asignación a los animales fue completamente aleatoria. Los datos son los siguientes:

T A B L A 6

Tiempos de supervivencia de los animales

Veneno	Tratamiento			
	A	B	C	D
I	0.31	0.82	0.43	0.45
	0.45	1.10	0.45	0.71
	0.46	0.88	0.63	0.66
	0.43	0.72	0.76	0.62
II	0.36	0.92	0.44	0.56
	0.29	0.61	0.35	1.02
	0.40	0.49	0.31	0.71
	0.23	1.24	0.40	0.38
III	0.22	0.30	0.23	0.30
	0.21	0.37	0.25	0.36
	0.18	0.38	0.24	0.31
	0.23	0.29	0.22	0.33

Los datos serán analizados mediante análisis de la varianza con una tabla de la siguiente forma:

Fuente	Grados de libertad	Suma de Cuadrados	Cuadrado Medio	Cociente
Venenos (p)	2	scp	cmp	$p = \text{cmp}/s^2$
Tratamientos (t)	3	sct	cmt	$t = \text{cmt}/s^2$
$p \times t$	6	sci	cmi	$i = \text{cmi}/s^2$
Dentro de los grupos (d)	36	scd	s^2	
Total corregido	47			

Supongamos que deseamos transformar los datos de forma que el valor esperado de la variable transformada en cada celda puede ser representado por medio de la suma de las constantes del renglón y la columna, es decir, que no se necesiten términos de interacción. Entonces desearíamos que los efectos de los medicamentos -cada veneno puede considerarse como un medicamento diferente- y de los tratamientos fueran grandes, y que los efectos de interacción fueran pequeños; de otra forma, nos gustaría que los cocientes p y t fueran relativamente grandes y el cociente i relativamente pequeño.

Consideramos la familia de transformaciones

$$W = \begin{cases} (y^\lambda - 1)/\lambda & , \lambda \neq 0 \\ \ln y & , \lambda = 0 \end{cases}$$

que sugieren Box y Cox. Para cualquier valor de λ podemos determinar los valores de p , t e i dados en la tabla. Como puede suceder que al aplicar algunas transformaciones las observaciones transformadas tengan varianzas distintas, sería útil examinar también alguna estadística que nos dé información sobre la homogeneidad de las variables transformadas. Describimos esta estadística a continuación (Levene, 1960):

Supongamos que se tienen p grupos de residuales, e_{ij} , como se indica:

$$\begin{aligned} \text{Grupo 1: } & e_{11}, e_{12}, \dots, e_{1n_1}, \text{ media } \bar{e}_1, \text{ Var } (e_{ij}) = \sigma_1^2 \\ \text{Grupo 2: } & e_{21}, e_{22}, \dots, e_{2n_2}, \text{ media } \bar{e}_2, \text{ Var } (e_{ij}) = \sigma_2^2 \\ & \vdots \\ \text{Grupo } p: & e_{p1}, e_{p2}, \dots, e_{pn_p}, \text{ media } \bar{e}_p, \text{ Var } (e_{ij}) = \sigma_p^2 \end{aligned}$$

De aquí, podemos construir las variables

$$z_{ij} = |e_{ij} - \bar{e}_i|, \quad j=1, 2, \dots, n_i, \quad i=1, 2, \dots, p$$

para obtener:

Grupo 1: $z_{11}, z_{12}, \dots, z_{1n_1}$, total Z_1

Grupo 2: $z_{21}, z_{22}, \dots, z_{2n_2}$, total Z_2

⋮

Grupo p: $z_{p1}, z_{p2}, \dots, z_{pn_p}$, total Z_p

Construimos ahora una tabla de análisis de la varianza como sigue:

Fuente | Suma de cuadrados | Grados de Libertad | Cuadrado Medio

Entre los grupos	$\sum_{i=1}^p \frac{z_i^2}{n_i} - \frac{G^2}{\sum n_i}$	$p-1$	s_1^2
------------------	---	-------	---------

Dentro de los grupos	por diferencia	$\sum_{i=1}^p (n_i-1)$	s^2
----------------------	----------------	------------------------	-------

Media	$\frac{G^2}{\sum n_i}$	1	
-------	------------------------	---	--

Total	$\sum_{i=1}^p \sum_{j=1}^{n_i} z_{ij}^2$	$\sum n_i$	
-------	--	------------	--

donde $G = Z_1 + Z_2 + \dots + Z_p$.

Obtenemos de aquí el valor $F_1 = S_1^2 / S^2$

Si $F_1 \geq F(p-1, \sum (n_i-1))$

diremos que hay eviden

cia de que existen diferencias entre $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$.

Si F_1 no es significativa, no rechazamos la hipótesis de que

$$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_p^2.$$

En este caso tenemos $p=12$ grupos del mismo tamaño en los cuales hay $n_i=4$, $i=1,2,\dots,12$, observaciones. Si obtenemos las cantidades de los cocientes p, t e i y de F_1 con intervalos para λ de 0.01 - unidades en el rango de - 2.0 a 1.2 resulta una gráfica como se muestra a continuación.

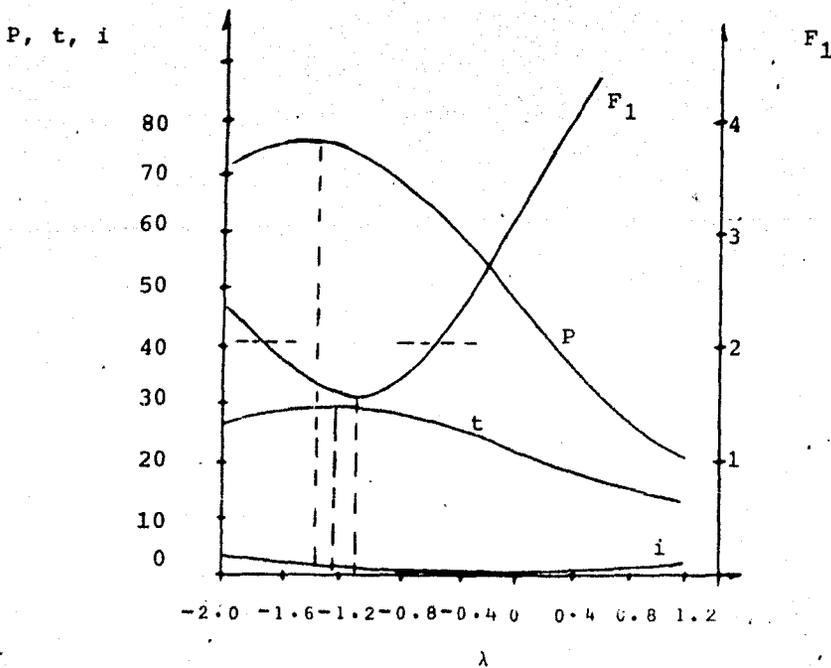


FIGURA 5

De esta gráfica podemos ver que el mínimo de i está aproximadamente en $\lambda = -0.60$, mientras que los mínimos de p y t se encuentran aproximadamente en $\lambda = -1.35$ y $\lambda = -1.25$, respectivamente. Si examinamos la gráfica de F_1 y usamos el valor crítico $F(11, 36) = 2.07$ -al .05 de significancia- como una base para la comparación, tendremos que restringirnos a valores de λ en el intervalo $(-1.75, -0.53)$, aproximadamente. Si le damos el peso principal al cociente i , el valor apropiado de λ sería -0.6 , aunque el valor de F_1 para $\lambda = -0.6$ nos haría reflexionar sobre tal elección. Sin embargo, como se ve la interacción no parece ser muy grande en ningún lado, mientras que la homogeneidad del error parece ser más importante aquí. De esta manera, podríamos concluir que el valor de $\lambda = -1$, que implicaría la transformación recíproca, puede ser un buen valor para λ (este mismo valor fue elegido por Box y Cox). Además, la transformación recíproca tiene aquí un especial atractivo: puede interpretarse como la tasa de mortalidad.

Si ajustamos un modelo simplificado (sin términos de interacción) a las observaciones transformadas, los residuales resultan de la siguiente manera:

T A B L A 7
Valores de los residuales

Veneno	Tratamiento			
	A	B	C	D
I	0.30	1.03	0.14	0.82
	-0.71	-0.28	0.03	0.01
	-0.76	-0.05	-0.60	0.12
	0.40	0.20	-0.87	0.21
II	-0.49	-0.44	-0.27	0.04
	0.18	0.11	0.33	-0.77
	-0.77	0.51	0.70	-0.34
	1.08	-0.72	-0.04	0.88
III	0.69	0.21	0.23	-0.01
	-0.10	-0.42	-0.13	-0.56
	0.70	-0.49	0.05	-0.11
	-0.51	0.33	0.43	-0.31

Aunque podríamos hacer un extenso estudio sobre estos residuales, sólo mencionaremos que parece que la única peculiaridad sucede en el tratamiento D en los factores I y III, ya que en el 1° todos son positivos y en el 2° todos son negativos. El modelo pues, no es totalmente satisfactorio para el tratamiento D, por lo cual sería conveniente tener mas información.

El método, como se ve en el ejemplo, nos provee de una gran ventaja: seleccionar la transformación asegurandonos de que posteriormente

se cumplirán algunos supuestos básicos requeridos. Además, como ya dijimos, puede utilizarse para transformar tanto la variable dependiente como la independiente simultáneamente, cumpliendo así con la demostración de Hill (1966) de que si ambas variables se han de transformar, debe hacerse al mismo tiempo y no en etapas diferentes.

CAPITULO

3

TRANSFORMACIONES ESPECIFICAS QUE MAS SE HAN UTILIZADO EN LA PRACTICA.

Históricamente se ha dado un énfasis mucho mayor a las transformaciones de la variable dependiente que a las de la variable independiente. Aunque esto no es muy claro en lo que se refiere a los métodos de transformación, si lo es, sin duda, en cuanto a las transformaciones específicas que se han estudiado: las de la variable dependiente son bastante numerosas. En este capítulo pretendemos hacer una recopilación de ellas y del uso que se les ha dado. Presentamos también los enunciados de una serie de proposiciones demostradas por Curtiss (1943), muy útiles en las transformaciones de la raíz cuadrada, la transformación angular y la logarítmica, cuando la variable tiene una distribución específica distinta de la Normal -el propósito es precisamente Normalizarla-. A lo largo de todo el capítulo la variable y denota la variable dependiente, $\sigma^2 = \text{Var}(y)$, $m = E(y)$ y $h(y)$ la función que define la transformación.

3.1 TRANSFORMACION POTENCIA.

Esta transformación representa en realidad a una familia de transformaciones; es la que consideran Box y Cox en su método:

$$h(y) = \begin{cases} \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1}, & \lambda_1 \neq 0 \\ \lim_{\lambda \rightarrow 0} \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} = \ln(y + \lambda_2), & \lambda_1 = 0 \end{cases}$$

válida para $y > -\lambda_2$ (de hecho la inclusión de λ_2 tiene el fin de lograr que $y + \lambda_2 > 0$ para toda y , como ya vimos). En la transformación más utilizada, con fines en ocasiones distintos que mencionaremos en sus casos particulares más importantes (3.2 a 3.5).

3.2 TRANSFORMACION RECIPROCA.

$$h(y) = y^{-1}$$

Podemos considerarla como la transformación adecuada para estabilizar la varianza, usando el método de Bartlett, cuando ésta tiende a ser proporcional a m^4 . En muchos casos es una transformación sugerida por los datos (ver el segundo ejemplo del método de Draper y Hunter) y en algunos de ellos cumple más satisfactoriamente los supuestos de Normalidad, varianza constante y aditividad. Casi siempre se usa cuando tiene un significado físico bien definido y cuando $P(y \leq 0)$ es despreciable.

3.3 TRANSFORMACION DE LA RAIZ CUADRADA.

$$h(y) \begin{cases} (y + \alpha)^{1/2}, & y \geq -\alpha \\ 0, & y < -\alpha \end{cases}$$

Consideramos tres casos:

a) Transformación de la raíz cuadrada para una variable con distribución de Poisson.

Cuando los datos estadísticos son números enteros, tales como el número de colonias de bacterias en una placa, el número de plantas en un área dada, etc., la variación de estos números y a menudo sigue la distribución de Poisson. Como para tal distribución $\sigma^2 = m$, con el método de Bartlett se obtiene fácilmente que para estabilizar la varianza debemos trabajar en la escala de la raíz cuadrada. Además, se ha demostrado la siguiente proposición:

Sea y una variable aleatoria que se distribuye Poisson con parámetro m . Si α es una constante arbitraria y si

$$T = f(y) = \begin{cases} (y + \alpha)^{1/2}, & y \geq -\alpha \\ 0, & y < -\alpha \end{cases}$$

entonces la distribución de $T - \sqrt{m+x}$ tiende a la distribución Normal cuando $m \rightarrow \infty$, con media cero y varianza $1/4$ (es decir, $\lim_{m \rightarrow \infty} \sigma_T^2 = 1/4$)

Bartlett (1947, p41) recomienda el uso de $\alpha = 1/2$ cuando las observaciones son números muy pequeños (i.e., para medias en el

rango de 10 a 2, especialmente cuando aparecen ceros entre los números observados). Posteriormente, Anscombe (1948) demostró que el valor de $\alpha = 3/8$ es óptimo para una estabilización de la varianza, con él obtenemos:

$$V(h) = \frac{1}{4} \left(1 + \frac{1}{16 m^2} \right)$$

En un trabajo más reciente Kihlberg, Herson y Schotz (1967) concluyen, después de un amplio estudio de computación, que $\alpha = 0.386$ -valor muy cercano al que da Anscombe- es óptimo para la mayoría de los valores de $E(y) = m$; aunque para valores pequeños de m , $m < 2$, el valor $\alpha = 1/4$ es mejor. Por último, Freeman y Tukey (1950), en un trabajo anterior al precedente sugieren el uso de la transformación.

$$h(y) = \sqrt{y} + \sqrt{y+1}$$

que funciona bastante bien para todos los valores de m .

En la práctica con frecuencia utilizamos un análisis de la varianza para datos de la forma mencionada -números enteros- porque sospechamos que se presentará heterogeneidad de algún modo, especialmente si los datos han sido recolectados bajo condiciones de campo. Entonces no es necesario suponer la distribución de Poisson, pero podemos usar aún la transformación de la raíz cuadrada si su variación parece estable.

- b) La transformación de la raíz cuadrada para una variable con distribución gamma

Aunque el caso anterior parece ser el más útil en la práctica, reproducimos aquí una proposición, demostrada también por Curtiss, que puede ser de algún provecho.

Sea y una variable aleatoria con una función de densidad del siguiente tipo:

$$F(y) = \begin{cases} 0, & \text{si } y \leq 0 \\ k y^{\frac{n}{2}-1} e^{-hy}, & \text{si } y > 0, h > 0 \end{cases}$$

donde $m = n/2h$.

Si α es una constante arbitraria y si

$$T = h(y) = \begin{cases} (y + \alpha)^{1/2}, & y \geq -\alpha \\ 0, & y < -\alpha \end{cases}$$

entonces la distribución de $T - \sqrt{(n/2h) + \alpha}$ tiende a la distribución Normal cuando $n \rightarrow \infty$, con media cero y varianza $1/4h$.

- c) Por último, se ha usado también la transformación $h(y) = (2y)^{1/2}$ para los casos en que y tiene una distribución χ^2 con v grados de libertad y Fisher (1925) ha demostrado que ésta h se distribuye aproximadamente Normal con media $\sqrt{2v-1}$ y varianza 1 (en adelante, en algunos casos, al hablar de esta distribución usaremos la notación $N(\sqrt{2v-1}, 1)$).

3.4 TRANSFORMACION DE LA RAIZ CUBICA

La aproximación de Fisher usando la raíz cuadrada no es del todo satisfactoria y Wilson y Hilferty (1931) sugirieron el uso de la transformación:

$$h(y) = (y/v)^{1/3}$$

que se distribuye aproximadamente $N(1 - \frac{2}{9v}, \frac{2}{9v})$ cuando y se distribuye χ^2 con v grados de libertad.

Haldane (1941) mostró que esta aproximación es adecuada para casi todos los propósitos prácticos y mejor que la de Fisher. Se ha encontrado también que esta transformación -al igual que la de la raíz cuadrada- puede ser útil al analizar datos sobre precipitación pluvial.

3.5 TRANSFORMACIONES LOGARITMICA E HIPERBOLICA

En algunos casos en que se presenta una heterogeneidad considerable en las observaciones, con frecuencia sucede que la varianza está correlacionada con la media, aún después de aplicar la transformación de la raíz cuadrada y podemos conseguir su estabilización aplicando una transformación logarítmica. La explicación natural de una varianza mayor que la media es que ésta misma varíe, es decir, que la varianza esté dada por una expresión del siguiente tipo:

$$\sigma^2 = m + \sigma_m^2$$

En algunas situaciones -p. ej. en poblaciones biológicas- esta variación de la media resulta proporcional a la misma media, de forma que podemos esperar que:

$$\sigma^2 = m + \lambda^2 m^2$$

expresión que, utilizando el método de Bartlett y con λ ó m grandes, implicaría una transformación logarítmica. En algunos problemas es posible que podamos estimar λ bastante bien para justificar una transformación más exacta que corresponda a una varianza de este tipo. Esta transformación puede ser:

$$h(y) = \lambda^{-1} \ln \left\{ \sqrt{1 + \lambda^2 y} + \lambda \sqrt{y} \right\} \quad \dots(1)$$

o, equivalentemente,

$$h(y) = \lambda^{-1} \operatorname{sen} h^{-1} (\lambda \sqrt{y}) \quad \dots(2)$$

Por ejemplo, se sabe que bajo ciertas suposiciones sobre la forma en que m varía, la distribución de Poisson se convierte en una distribución binomial negativa. Para tales datos, la transformación (2) puede ser apropiada. Esta transformación tiene la desventaja de requerir un conocimiento aproximado de λ , y la transformación $h(y) = \ln(y + c)$ parece resultar bastante buena en muchos casos ⁷.

⁷ La inclusión de una constante c se hace para evitar las dificultades de los ceros.

La transformación $h(y) = \lambda^{-1} \operatorname{sen} h^{-1}(\lambda\sqrt{y})$ también es apropiada cuando tenemos una varianza del tipo más general

$$\sigma^2 = \mu^2 (m + \lambda^2 m^2).$$

En algunos casos, particularmente en estudios biológicos y de crecimiento de una población, se ha visto que la probabilidad de ocurrencia de un evento, y , puede expresarse como $P(y) = 1/[1 + \exp -(a+by)]$. En situaciones de este tipo, ha resultado ser útil la transformación "logit": $h(y) = \log (y/1-y)$.

Aunque no es muy común tener que analizar un conjunto de coeficientes de correlación mediante análisis de la varianza, hay una transformación que puede ser apropiada si el problema se presenta. Esta transformación es $1/2 \ln \left\{ \frac{1+r}{1-r} \right\}$, donde r es el coeficiente de correlación muestral. El propósito de la transformación es conseguir que la distribución de r tenga una desviación menor y una varianza más estable, cuando deseamos que la varianza del verdadero coeficiente -el de la población-, ρ , sea independiente del mismo.

Un problema más importante y que ocurre con frecuencia, es el análisis de la varianza de un conjunto de varianzas o desviaciones estándar de una muestra. La varianza del estimador de la varianza poblacional, s^2 , es proporcional a $(\sigma^2)^2$, por tanto, usando de nuevo el método de Bartlett, la transformación logarítmica puede ser

adecuada en este caso.

Curtiss ha demostrado tres proposiciones sobre transformaciones logarítmicas muy relacionadas con las mencionadas en este apartado. Las proposiciones son las siguientes:

Sea y una variable aleatoria con distribución de frecuencias relativas binomial con parámetro p y los n valores $0, 1/n, 2/n, \dots, n/n$.

$$1) \quad \text{Si } T=h(y) = \begin{cases} \sqrt{n} \ln (\sqrt{y} + \sqrt{y+1}) = \sqrt{n} \operatorname{sen}^{-1} \sqrt{y}, & y \geq 0 \\ 0 & , \quad y < 0 \end{cases}$$

entonces la distribución de $T - \sqrt{n} \operatorname{sen}^{-1} \sqrt{p}$ tiende a la distribución Normal cuando $n \rightarrow \infty$, con media cero y varianza $(1-p)/(4+4p)$.

$$2) \quad \text{Si } T=h(y) = \begin{cases} \sqrt{n} \ln y, & y > 0 \\ 0 & , \quad y \leq 0 \end{cases}$$

entonces la distribución de $T - \sqrt{n} \ln p$ tiende a la distribución Normal cuando $n \rightarrow \infty$, con media cero y varianza $(1-p)/p$.

$$3) \quad \text{Si } T=h(y) = \begin{cases} \frac{1}{2} \sqrt{n} \log (y/1-y), & 0 < y < 1 \\ 0 & , \quad y \leq 0, y \geq 1 \end{cases}$$

entonces la distribución de $T - \frac{1}{2} \sqrt{n} \ln (p/1-p)$ tiende a la distribución Normal cuando $n \rightarrow \infty$, con media cero y varianza $1/4p(1-p)$.

Además de estas consideraciones que están basadas fundamentalmente en la aplicación del método de Bartlett, hay otro aspecto

importante sobre la transformación logarítmica: existe una distribución específica, la distribución log normal, que ha sido ampliamente estudiada y que representa muy bien muchas situaciones (p. ej. el diámetro de partículas pequeñas sometidas a un proceso de compresión, la vida promedio de algunos artículos, la distribución de los ingresos de una población, etc.) En este caso se considera la transformación ya vista

$$h(y) = \ln(y + c)$$

de manera que si (y solo si) h es Normal entonces y se distribuye log normalmente ⁸.

3.6 TRANSFORMACION ANGULAR

Curtiss, una vez más, ha demostrado una proposición que hace uso de esta transformación. La enunciamos a continuación:

Sea y una variable aleatoria con una distribución de frecuencias relativas binominal con parámetro p y los n valores 0, 1/n, 2/n, ..., n/n. Si α es una constante arbitraria y si

$$T = h(y) = \begin{cases} \sqrt{n} \operatorname{sen}^{-1} \sqrt{y + (\alpha/n)}, & -\alpha/n \leq y \leq 1 - \alpha/n \\ 0 & , y < -\alpha/n, y > 1 - \alpha/n \end{cases}$$

donde T es medida en radianes, entonces la distribución de

$T - \sqrt{n} \operatorname{sen}^{-1} \sqrt{p + (\alpha/n)}$ tiende a la distribución Normal cuando $n \rightarrow \infty$, con media cero y varianza 1/4.

⁸ De otra forma, supongamos que h se distribuye $N(\mu, \sigma_h^2)$, entonces $y = e^h$ tiene una distribución lognormal.

Bartlett (1936) da unos resultados numéricos en los casos $n=10, \alpha=0$ y $n=10, \alpha=1/2$, los cuales indican que tal vez la elección de $\alpha=1/2$ es más adecuada si la p estimada está cerca de 0 ó 1 y $\alpha=0$ es preferible si ésta se encuentra entre 0.3 y 0.7.

Posteriormente se ha considerado esta transformación como:

$$h(y) = \text{sen}^{-1} \left(\frac{y + \alpha_1}{n + \alpha_2} \right)^{1/2}$$

en la que Bartlett sugiere el uso de $\alpha_1 = 1/2$ y $\alpha_2 = 0$ y después Anscombe (1948) recomienda los valores de $\alpha_1 = 3/8$ y $\alpha_2 = 3/4$, multiplicando toda la expresión por el factor $\sqrt{\frac{n+1}{2}}$. Estas últimas se han utilizado principalmente para intentar estabilizar la varianza.

3.7 TRANSFORMACION LOGARITMICA DOBLE

En muchos fenómenos físicos y biológicos, podemos expresar la probabilidad de un cambio como $p = 1 - \exp(-by)$. En tales casos, Fisher y Yates (1957) han demostrado que la transformación

$$h(y) = \ln(-\ln y), \quad 0 < y < 1$$

es una transformación útil y con efectos muy similares a otras transformaciones como la angular y la de probit. Se ha utilizado también en las estadísticas de extremos donde la distribución acumulada asintótica del menor valor es:

$$F(y) = 1 - \exp(-e^{-x})$$

con $x = \alpha_n (y - y_{(n)})$, donde $y_{(n)}$ es el valor más grande y α_n un factor con dimensión y^{-1} .

3.8 TRANSFORMACION INTEGRAL DE PROBABILIDAD

Pearson (1938) ha considerado la transformación

$$h(y) = \int_{-\infty}^y p(t) dt$$

donde $p(y)$ es la función de densidad de y . En este caso, h se distribuye uniformemente en el intervalo $(0,1)$, de manera que esta transformación no conduce a la distribución Normal pero estabiliza perfectamente la varianza en $V(h) = 1/12$.

3.9 TRANSFORMACION COORDENADA

Otra transformación que se ha utilizado es

$$h(y) = N^{-1} \{ F(y) \}$$

donde $F(y)$ es la función de distribución acumulada de y y N^{-1} es la inversa de la función de distribución de la distribución Normal

estándar. En este caso h es exactamente una variable Normal estándar pero, por supuesto, en general desconocemos F y no podemos expresar N^{-1} en una forma exacta. Sin embargo, Kowalski y Tarter (1969) demostraron que usando aproximaciones a N^{-1} y el estimador de Fourier de la función de distribución, es posible utilizar la transformación en la práctica (si se desea conocer este estimador se pueden ver en el artículo en cuestión).

3.10 TRANSFORMACION PROBIT

Es una transformación que se define mediante la función inversa de h de la siguiente manera:

$$y = \int_{-\infty}^{h-5} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad 0 < y < 1$$

Esto se hace así ya que no existe la primitiva de la función de distribución Normal, pero Finney (1964) obtuvo, mediante aproximaciones, unas tablas de los valores de h para distintos valores de y . Se considera el valor $h-5$ con el objeto de hacer muy rara la ocurrencia de valores negativos. Se ha utilizado principalmente en estudios biológicos en los que la variable dependiente está constituida por las proporciones de muertes en una población al aplicarles ciertas dosis (variable independiente). Usualmente con ella se pretende obtener una relación lineal, por medio de transformaciones tanto a las proporciones de la población como a las dosis aplicadas (para esta últi

ma variable se ha utilizado la transformación logarítmica con el fin de Normalizarla).

Actualmente está en preparación una tesis de Actuaría que presenta un estudio bastante completo sobre esta transformación y sus aplicaciones.

CAPITULO

4

PROCEDIMIENTOS PARA VERIFICAR LA UTILIDAD DE LA TRANSFORMACION

Siempre que se haya utilizado una transformación es sumamente importante analizar las propiedades del modelo obtenido: si es posible presentar los resultados en la variable original sin afectar las estimaciones, si las variables transformadas realmente explican mejor el comportamiento de las observaciones, en resumen, si hemos alcanzado o no los objetivos de la transformación. Presentamos en este capítulo algunos procedimientos utilizados para este fin.

4.1 OBTENCION DEL SESGO DE LA TRANSFORMACION

Cualquiera que sea el propósito de la transformación, con frecuencia surgen problemas cuando ya hemos hecho el análisis de los datos transformados. Por ejemplo, en los experimentos sobre la modificación del clima, como ya hemos dicho, el análisis de los datos puede ser más efectivo en términos de raíz cúbica de la cantidad de lluvia y servirá para indicar cuándo se presenta algún efecto debido a una alteración artificial. Supongamos que la evidencia sugiere que así es. Obviamente, la magnitud estimada del efecto no puede presentarse en función de la raíz cúbica de pulgadas de lluvia y no sabemos si será suficiente tan sólo elevar al cubo el efecto tal y como es medido en las variables transformadas y presentarlo como

el estimador insesgado de mínima varianza del efecto de las alteraciones artificiales.

De manera más general, tenemos las transformaciones $h(y)$, los estimadores $\hat{m}_h = \hat{E}(h)$ y $\hat{\sigma}_h^2 = \hat{V}\text{ar}(h)$ y conocemos su distribución. El problema es encontrar el estimador insesgado de varianza mínima (EIVM) de $m(m=E(y))$, \hat{m} . En la práctica, desearíamos también tener $\text{Var}(\hat{m})$ y un estimador de ésta, $\hat{V}\text{ar}(\hat{m})$, y, utilizando éstos, ser capaces de dar un intervalo de confianza para m .

Neyman y Scott (1960) han considerado este problema en términos generales. Han demostrado que si $f(h)$ (si $h=g(y)$ es la transformación, entonces la función inversa, h^{-1} , es $y=f(h)$) es cualquier función entera de segundo orden ⁹ o menos, entonces m y el EIVM de m existen.

Si $f(z)$ es, pues, una función entera de segundo orden, entonces:

$$m = f(0) + \sum_{r=1}^{\infty} \frac{1}{r!} f_0^{(r)} E(h^r)$$

y,

$$\hat{m} = \hat{m}(\hat{m}_h, s) = f(0) + \sum_{r=1}^{\infty} \frac{1}{r!} f_0^{(r)} \text{Tr},$$

donde s/V es el EIVM de σ_h^2 (y tal que $s/\sigma_h^2 \sim \chi^2_{(v)}$), $f_0^{(n)}$ representa la n -ésima derivada de f valuada en a y Tr es tal que, $E(\text{Tr}) = E(h^r)$, en concreto:

9. Ver apéndice

$$T_{2r} = \sum_{k=0}^r \frac{(2r)!}{(2k)! (r-k)!} \hat{m}_h^{2k} \left[\frac{1}{4} S(1-\lambda^2) \right]^{r-k} \frac{\Gamma(v/2)}{\Gamma\left(\frac{v}{2} + r-k\right)},$$

$$Y, \quad T_{2r+1} = \sum_{k=0}^r \frac{(2r+1)!}{(2k+1)! (r-k)!} \hat{m}_h^{2k} \left[\frac{1}{4} S(1-\lambda^2) \right]^{r-k} \frac{\Gamma(v/2)}{\Gamma\left(\frac{v}{2} + r-k\right)}$$

donde λ^2 es tal que el EIVM de m_n , \hat{m}_h , se distribuye $N(m_h, \lambda^2 \sigma^2)$, es decir, λ^2 define la varianza de m_h como un múltiplo de la varianza de la variable aleatoria h .

El \hat{m} obtenido de esta forma es el EIVM de m . Además como \hat{m}_h y σ_h^2 son estadísticas suficientes y completas, \hat{m} es único. Aunque esta fórmula es algo complicada, señala que 4 transformaciones usadas comúnmente -la raíz cuadrada, la logarítmica, la angular y la hiperbólica- están relacionadas por una ecuación diferencial simple que involucra a sus inversas. Estas funciones también son funciones enteras. La ecuación es:

$$f_h^{(2)} = A + Bf(h)$$

La siguiente tabla muestra los valores de A y B para estas 4 transformaciones:

T A B L A 8

Transformaciones recursivas

$h=g(y)$	$y=f(h)$	A	B
$y^{1/2}$	h^2	2	0
$\log_n y$	$\exp(\ln N) = e^n$	0	n^2
$\text{sen}^{-1} \sqrt{y}$	$\text{sen}^2 h$	2	-4
$\text{sen} h^{-1} \sqrt{y}$	$\text{sen} h^2 h$	2	4

Si hacemos la transformación después de que una función lineal ha sido ajustada a la variable original, no surgen problemas adicionales. Sea z la variable original, con media m_z y sea $y = cz + d$ ($c \neq 0$). Entonces $m = E(y) = cm_z + d$, y $\hat{m}_z = \frac{1}{c} (\hat{m} - d)$. \hat{m}_z es el único EIVM de m_z .

Neyman y Scott obtienen los siguientes resultados para m y

\hat{m} :

$$m = \begin{cases} f(\hat{m}_h) \exp(B\sigma_h^2/2) + \frac{A}{B} [\exp(B\sigma_h^2/2) - 1], & B \neq 0 \\ f(\hat{m}_h) + A\sigma_h^2/2, & B = 0 \end{cases}$$

$$\hat{m} = \begin{cases} \phi[B(1-\lambda^2)S, v] \left[\frac{f(\hat{m}_h) + \frac{A}{B}}{B} \right] - \frac{A}{B}, & B \neq 0 \\ f(\hat{m}_h) + A(1-\lambda^2)\sigma_h^2/2, & B = 0 \end{cases}$$

$$\begin{aligned} \text{donde } \phi(as, v) &= \sum_{r=0}^{\infty} \frac{1}{r!} \frac{\Gamma(v/2)}{\Gamma(\frac{v}{2} + r)} \left(\frac{as}{4}\right)^r \\ &= \left(\frac{2}{S\sqrt{a}}\right)^{v/2-1} \Gamma(v/2) I_{\frac{v}{2}-1}(S\sqrt{a}) \end{aligned}$$

10

$In(t)$ es la función de Bessel de argumento imaginario. Esta serie converge muy rápidamente, usualmente sólo se requieren pocos términos para conseguir una precisión adecuada.

De estos resultados, determinamos el sesgo del estimador

10 La función de Bessel de 1a. clase es la función definida por la serie:

$$In(t) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k! (n+k)!} \left(\frac{t}{2}\right)^{n+2k}, \quad n=0,1,2,\dots$$

$f(\hat{m}_h)$ de m :

$$E[f(\hat{m}_h) - m] = \begin{cases} \left[\frac{m+A}{B} \right] [\exp \{ -B(1-\lambda^2) \sigma_h^2 / 2 \} - 1] , & B \neq 0 \\ -A(1-\lambda^2) \sigma_h^2 / 2 , & B = 0 \end{cases}$$

Es posible demostrar que el valor absoluto de este sesgo siempre es una función monótona decreciente de λ^2 . Esto implica, recordando que λ^2 mide la precisión -en términos de σ_h^2 - con la que m_h es estimada, que mientras mayor sea el tamaño de muestra ($\lambda^2 = 1/n$) o mejor sea el diseño experimental, es peor el sesgo del estimador $f(\hat{m}_h)$. Por tanto, si el análisis de los datos es incorrecto y no hacemos ninguna corrección para permitir el sesgo en la transformación inversa, algunas de las ventajas de una muestra grande o de un diseño eficiente se perderán.

Hay muchas transformaciones cuya función inversa es una función entera pero que no es una solución a la condición recursiva de Neyman y Scott. En estos casos, normalmente deben usarse los resultados más generales. Para algunas transformaciones sencillas, como por ejemplo la raíz cúbica, podemos obtener fácilmente una expresión para \hat{m} mediante el manejo de las expresiones conocidas para los momentos de las distribuciones Normal y ji-cuadrada.

Es interesante notar que los resultados indican una dificultad teórica con la transformación recíproca, $h(y) = 1/y$. En este caso, $m = E(y) = E(1/h)$ no existe, de manera que no podemos estimarla. Sin

embargo, Box (1971) ha sugerido la definición de la pseudo-esperanza de y , m' , como:

$$m' = PE(y) = \lim_{\epsilon \rightarrow 0} \int_{\epsilon}^{\infty} \frac{1}{h} E^{-1} \left[\frac{e^{-\frac{(h-m_h)^2}{2\sigma_h^2}}}{\sqrt{2\pi}\sigma_h} \right] dh$$

donde E^{-1} es el EIVM de la expresión entre llaves.

4.2 UTILIZACION DEL COEFICIENTE DE CORRELACION MULTIPLE, R^2 , PARA VERIFICAR LA TRANSFORMACION

Es un procedimiento que, por medio de la comparación de los coeficientes de determinación (o correlación) para la variable original y la transformada, permite afirmar que el mejor modelo es el de la variable original cuando la transformación cumple con ciertas hipótesis. El método se basa en el siguiente teorema:

Sea y una variable aleatoria dependiente de un conjunto de k variables aleatorias $\underline{x} = (x_1, x_2, \dots, x_k)$. Supongamos que existe una función $f(\underline{x})$ tal que (hipótesis básicas):

$$y = f(\underline{x}) + e$$

donde $f(\underline{x})$ y e se distribuyen Normalmente y e es independiente de \underline{x} , con $E(e) = 0$ y $E(e^2) > 0$.

Ahora, sea z alguna variable aleatoria y $g_z(\underline{x})$ una función de \underline{x} seleccionada de forma que la varianza de

$$e_z = z - g_z(\underline{x})$$

sea tan pequeña como sea posible. Podemos definir entonces la can-
tidad (de la población)

$$R_z^2 = 1 - \frac{\text{Var}(e_z)}{\text{Var}(z)}$$

Se considera una clase de transformaciones, $h(y, \theta)$, dependiendo del parámetro θ -que puede ser un vector-, tal que:

$$T(y, \theta_0) = y$$

y con la propiedad de que

$$T(y, \theta) = \sum_{j=0}^m \alpha_j(\theta) H_j(y)$$

donde m no necesita ser finita y $H_j(y)$ es el j -ésimo polinomio de Hermite. Esta restricción de que una transformación pueda expand
se en términos de los polinomios de Hermite no es muy fuerte (ver demostración).

Entonces, si $z = T(y, \theta)$, con $\theta \neq \theta_0$ se tiene

$$R_z^2 < R_y^2$$

El teorema, entonces, demuestra que el grado en que y explica la variación de los datos -medido en términos de R^2 - (o de otra forma, el grado de explicación de y obtenido de \underline{x}) es mayor que el

de cualquier transformación de y , si se cumplen las hipótesis básicas.

Antes de dar la demostración del teorema, veremos algunas implicaciones importantes. En la construcción de modelos económicos, con frecuencia la teoría no es lo suficientemente precisa en la especificación de las formas de la función. En tales circunstancias podríamos ajustar varias ecuaciones del tipo $f(y) = g(x) + \text{error}$, donde f y g son dos funciones que pensamos que posiblemente sean adecuadas.

En muchas ocasiones se piensa que, en esta situación, la comparación directa de los valores de R^2 no es significativa. Sin embargo, el teorema anterior sugiere otra cosa cuando se cumplen las hipótesis. Supongamos que un modelo específico cumple los supuestos fundamentales de la formulación (1). En tales circunstancias sería razonable considerar que ése es el modelo correcto. El teorema afirma que el modelo correcto es aquél para el cual R^2 es mayor. Con esta formulación, la función no lineal óptima de x para explicar cualquier otra transformación de y puede obtenerse de la ecuación (3) que damos en la demostración.

Debemos tener en cuenta, sin embargo, las siguientes dificultades prácticas:

- a) El teorema es un resultado acerca de cantidades poblacionales y no acerca de estadísticas derivadas de una muestra finita. Entonces este criterio no servirá para elegir invariablemente el mode-

lo correcto. No obstante, sí lo hará asintóticamente, pues R^2 muestral es consistente para el valor de la población.

- b) Podemos considerar fácilmente ejemplos en que ninguna transformación de x y y , de la forma mencionada, satisface los supuestos básicos. En tales situaciones el teorema no es aplicable, aún a cantidades de la población. Por ejemplo, supongamos que el verdadero modelo es

$$y = a + bx + e$$

donde la distribución de e no es Normal y es independiente de x . En este caso la aplicación del teorema no necesariamente conseguirá la forma funcional correcta. Algunos trabajos empíricos sobre transformaciones suponen que la relación funcional correcta producirá errores distribuidos Normalmente, y si esta suposición es falsa, los métodos empleados no son estrictamente válidos.

- c) Al considerar una clase particular de transformaciones, como la de Box y Cox:

$$z = h(y, \lambda) = \frac{y^\lambda - 1}{\lambda}$$

el estimador de λ obtenido aplicando el teorema a z para varios valores de λ , para una muestra dada, no necesariamente coincide con el obtenido con un desarrollo por máxima verosimilitud.

Demostración del Teorema

El polinomio de Hermite de n -ésimo orden está definido por:

$$\begin{aligned} H_n(x) &= \exp\left[\frac{x^2}{2}\right] (-d/dx)^n \exp\left[-\frac{x^2}{2}\right] \\ &= n! \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{(-1)^m}{m!} (2^m m! (n-2m)!)^{-1} x^{n-2m} \end{aligned}$$

donde $[n/2]$ es el mayor entero $\leq n/2$.

Sea E_0 el operador definido por:

$$E_0 [\psi(x)] = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \psi(x) e^{-x^2/2} dx$$

es decir $E_0 [\psi(x)]$ es la esperanza de la transformación $\psi(x)$ de x , donde $x \sim N(0,1)$.

Los polinomios de Hermite tiene las siguientes propiedades:

$$\begin{aligned} E_0 [H_n(x) H_k(x)] &= n! \quad , \quad n = k \\ &= 0 \quad , \quad n \neq k \end{aligned}$$

y como $H_0(x) = 1$, se sigue que $E_0 [H_n(x)] = 0, n > 0$

Si x y y tienen una distribución conjunta según la $N(0,1)$ y con coeficiente de correlación ρ , entonces:

$$\begin{aligned} E_0 [H_n(x) H_k(y)] &= \rho^n n! \quad , \quad n=k \\ &= 0 \quad , \quad n \neq k \end{aligned}$$

donde la esperanza es sobre la distribución Normal bivariada. Podemos representar una función $T(x)$ mediante una expansión en series de los polinomios de Hermite, en la forma:

$$T(x) = \sum_{j=0}^{\infty} \alpha_j H_j(x)$$

donde $\alpha_k = E_0 [T(x) H_k(x)]$, si se cumple que $E_0 [T(x)] < \infty$

Suponiendo que $A^2 + B^2 = 1$,

$$H_n(Ax+By) = \sum_{k=0}^n n c_k A^k B^{n-k} H_k(x) H_{n-k}(y) \quad \dots (2)$$

Sean $E(x) = \mu$ (entonces $E(f(x)) = \mu$), $V(y) = \sigma^2$, $V(f(x)) = \sigma_f^2$ y

$V(e) = \sigma_e^2$. Entonces, bajo las hipótesis del teorema:

$$\sigma^2 = \sigma_f^2 + \sigma_e^2$$

y la probabilidad condicional de y dado x es $N(f(x), \sigma^2)$. Por tanto, al definir

$$z = T \left(\frac{y - \mu}{\sigma} \right) = \sum_{j=1}^m \alpha_j H_j \left(\frac{y - \mu}{\sigma} \right)$$

deducimos, por las propiedades de los polinomios de Hermite, que la media y la varianza de z es:

$$E_0(z) = \alpha_0 \quad \text{y} \quad E_0[(z - \alpha_0)^2] = \sum_{j=1}^m j! \alpha_j^2$$

Ahora, para encontrar la media condicional de z dado x , necesitamos usar la descomposición:

$$\frac{y - \mu}{\sigma} = AP + BQ,$$

$$\text{donde} \quad A = (1 - \sigma_e^2 / \sigma^2)^{1/2}, \quad P = (f - \mu) / \sigma_f$$

$$B = \sigma_e / \sigma, \quad Q = e / \sigma_e$$

Expandiendo cada término $H_j [(y - \mu) / \sigma]$ mediante (2) e integrando obtenemos que:

$$E_0(z/x) = g_z(x) = \sum_{j=0}^m \alpha_j A^j H_j(P) \quad \dots (3)$$

entonces $g_z(x)$ será una variable aleatoria y, dado que $P \sim N(0,1)$, su media y su varianza serán:

$$E_0[g_z(x)] = \alpha_0 \quad \text{y} \quad \text{Var}[g_z(x)] = \sum_{j=0}^m j! A^{2j} \alpha_j^2$$

Por tanto, tenemos que

$$R_z^2 = \frac{\sum_{j=1}^m j! A^{2j} \alpha_j^2}{\sum_{j=1}^m j! \alpha_j^2}$$

y como $0 < A^2 < 1$ y $R_y^2 = A^2$ se sigue que

$$R_z^2 < R_y^2$$

siempre que $m > 1$.

La igualdad $R_z^2 = R_y^2$ se daría sólo si $m=1$ ó si $A^2 = 1$, casos que hemos supuesto que no se dan. La condición $m=1$ implicaría que hemos usado una transformación lineal y $A^2=1$ implicaría que la ecuación $y=f(x) + e$ es ajustada con una varianza en el error igual a cero.

4.3 UN PROCEDIMIENTO PARA AVERIGUAR LOS EFECTOS DE LA TRANSFORMACION

Lindsey (1975), que ha sugerido este método, supone que la transformación puede llevar a cabo tres funciones distintas:

- (1) Conseguir que la variable dependiente cumpla mejor la hipótesis de que todos los cumulantes después del segundo son cero ¹¹

¹¹ Esta es una característica determinante de la distribución Normal, de forma que este primer punto podría escribirse: conseguir que la variable dependiente tenga una distribución Normal, al menos aproximada.

- (2) Hacer que la varianza esté más próxima a ser constante entre los puntos en el espacio de los factores (espacio x)
- (3) Cuando el modelo matemático, que describe la variación de la media entre los puntos, contiene menos parámetros que puntos observados en el espacio (i.e. cuando se tiene una ecuación de regresión o un análisis de varianza sin interacción), conseguir que este modelo explique mejor la variación observada.

El estimador de máxima verosimilitud (MV) sirve como una clase de promedio al llevar a cabo estas 3 funciones. Pero es posible distinguir cada una de ellas usando la función de verosimilitud. Hay que hacer notar dos cosas: la misma transformación se supone para conseguir que la función (1) se cumpla en todos los puntos del espacio x . Si se dispusiera de un número grande de observaciones en cada punto, podría estimarse un valor distinto de la transformación en cada punto observado. En segundo lugar, puede ser que los 3 objetivos no sean aplicables en una situación dada.

Sea γ un vector de parámetros desconocidos que podemos estimar, usado para transformar las variables dependientes observadas de alguna manera específica para una distribución de probabilidad dada. La variable dependiente, y , se convierte en la variable transformada $h(y, \gamma)$, donde h es la función de transformación específica. Al aplicar tal transformación, debemos incluir el ja-

cobiano adecuado en la función de probabilidad.

Consideremos ahora las tres funciones individualmente. Su pongamos que una media diferente y una varianza diferente son ajustadas para cada punto. Por supuesto, para poder hacer esto, al menos debemos disponer de 3 observaciones en cada punto. Entonces podemos obtener el estimador de MV, $\hat{\gamma}_1$, del vector de transformación. Si γ_1 es de una o dos dimensiones, podemos graficar la función de verosimilitud maximizada con respecto a todas las medias y varianzas. Con esta distribución de probabilidad, sólo se ve involucrada la función (1) puesto que la varianza no es constante y hay tantos parámetros para la media como puntos. La superficie de la función de verosimilitud para γ_1 será usualmente muy plana, mostrando un rango muy amplio para los valores factibles de los parámetros, a menos que se cuente con un gran número de observaciones en cada punto.

Supongamos ahora que utilizamos la misma distribución pero con varianza constante y la transformación común γ_2 , para todos los puntos que tienen la misma forma que γ_1 . De nuevo, el estimador de MV se puede obtener y se inspecciona la superficie de la función de verosimilitud. En este caso, ambas funciones, (1) y (2), están involucradas. En general, la superficie de la función de verosimilitud para γ_2 será considerablemente diferente de la de γ_1 .

El tercer caso es el considerado normalmente: la estimación de la transformación, γ_3 , para el modelo completo con varianza constante e incluyendo el modelo matemático requerido para la media. La superficie de la función de verosimilitud para γ_3 podrá o no diferir de la de γ_2 .

El mismo Lindsey da un ejemplo en el que ajusta una variable dependiente de dos factores, que consiste en una ecuación de regresión no lineal con 8 parámetros, a datos que representan las proporciones de huevos de pescado en cría. Fueron tomadas 4 observaciones en cada uno de los 18 puntos del espacio x. Se usó la transformación potencia de Box y Cox, y^Y para obtener un mejor ajuste. Cuando se aplica el análisis precedente, las gráficas de las funciones resultan como se muestra en la figura 6. Con estos datos, la transformación cumple primariamente la 2a. y 3a. función. Una vez que hemos cumplido los supuestos de la ecuación de regresión y de varianza constante, aún permanece una pequeña indicación de que los datos no son Normales. Con $\hat{\gamma}_2=0.48$, podemos considerar $y^{0.48}$ como una variable Normal con varianza constante sobre los 18 puntos que intervinieron en la muestra. Con $\hat{\gamma}_3 = 0.84$, podemos considerar $y^{0.84}$ como una variable Normal con varianza constante y media variando de acuerdo a la ecuación de regresión estimada. Entonces, $\gamma_4 = \gamma_3 / \gamma_2 = 1.75$ transforma la 1a. de estas variables en la segunda: $(y^{\gamma_2})^{\gamma_4} = y^{\gamma_3}$. Por tanto, γ_4 es la parte de la transformación γ_3 que cumple con la función (3). Similarmente, $\gamma_5 = \gamma_2 / \gamma_1$ nos da la parte de la transfor

mación que cumple con la función (2). Por supuesto, γ_1 es la parte que cumple con (1). Entonces la transformación se descompone en $\gamma_3 = \gamma_5 \gamma_4 \gamma_1$. Esta descomposición de γ_3 es un rasgo distintivo de la transformación potencia y, en general, no se podrá hacer siempre así.

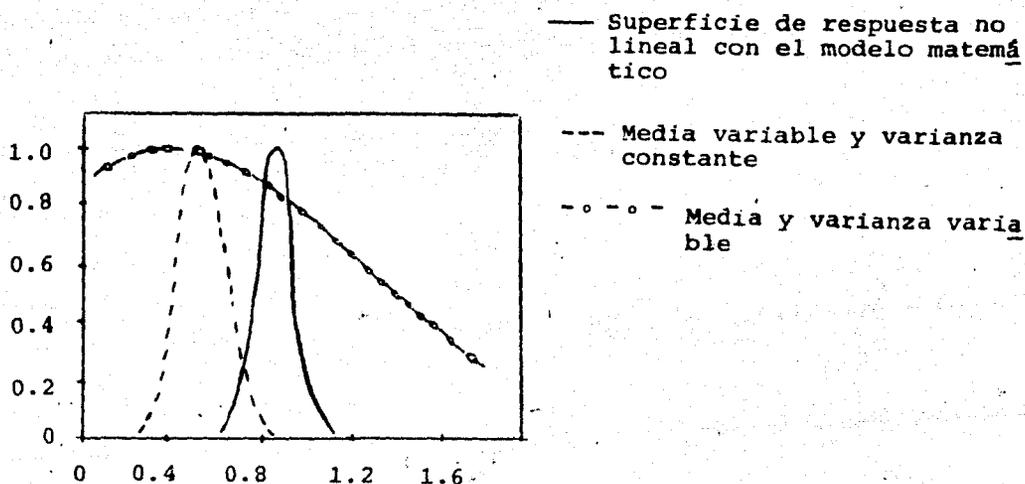


FIGURA 6

El procedimiento también fue aplicado al primer ejemplo dado por Draper y Hunter (ver el 2º ejemplo del método). Se cuenta en ese caso con 4 observaciones en cada uno de los 12 puntos y se supone que no hay iteración. La gráfica de la función de verosimilitud para γ_1 es relativamente plana -como en la fig. 6 para el ejemplo anterior- mientras que las de γ_2 y γ_3 son prácticamente idénticas. Esto indica que una vez que hemos transformado la variable para cumplir las primeras dos funciones, la función (3) se cumple automáticamente. En

otras palabras, el objetivo principal que logramos con la transformación potencia para estos datos, es hacer la varianza relativamente constante, mientras que la eliminación de los efectos de interacción depende de aquél.

El orden lógico de los objetivos de la transformación puede modificarse intercambiando (2) y (3). En este caso, estimamos γ_2 para la distribución con una varianza distinta en cada punto pero usando la hipótesis del modelo matemático de la media. Los parámetros γ_1 y γ_3 permanecen iguales.

En resumen, si consideramos que los 3 aspectos mencionados al principio son los objetivos de cualquier transformación, la importancia relativa de cada uno de ellos puede ser determinada por el análisis de las superficies de verosimilitud resultantes.

C A P I T U L O 5

CONCLUSIONES

Como ya hemos dicho, antes de precipitarse hacia el uso de una transformación, debemos atender a una serie de aspectos que pueden llevarnos a tomar la decisión de que una transformación es inconveniente. Antes que nada hemos de tener fijos los objetivos del análisis que deseamos hacer. Una vez hecho esto y si nos parece que el transformar puede tener sentido, debemos proceder a analizar los aspectos mencionados en el primer capítulo. Si aún después de esto vemos que el uso de una transformación puede ser de alguna utilidad, es importante hacer primero un análisis de los datos ya que éstos pueden sugerir alguna transformación específica (ver Draper y Hunter, 1969, ejemplo 1), antes que aplicar algún método o transformación. Los objetivos de la transformación -que se supone hemos fijado previamente- pueden constituir una ayuda bastante eficaz en nuestra decisión sobre el método o la transformación que debemos utilizar. Nos parece importante también destacar la recomendación de Draper y Hunter: puede resultar muy útil el análisis simultáneo de la aplicación de criterios o transformaciones distintas, puesto que nos permite tomar una postura flexible siempre que no nos inclinamos por una transformación específica en el inicio (es posible olvidarse momentáneamente de aquéllas sugeridas por los datos), aunque esto supondrá, en la mayoría de los casos, la ayuda de una computadora.

Cuando se pretende transformar tanto la variable dependiente como la independiente, Hill (1966) ha mostrado que todas las transformaciones deben llevarse a cabo simultáneamente y no en etapas.

Por otra parte, es también importante la verificación de las hipótesis una vez que se ha aplicado alguna transformación. De la misma manera que las hemos verificado antes de aplicarla, hemos de hacerlo después ya que puede haber situaciones en que la transformación no haya conseguido alguno (o incluso ninguno) de los objetivos o, lo que es peor, puede suceder que la transformación consiga alguna de las hipótesis pero destruya otras. No obstante, siempre hay que tener en cuenta que los procedimientos usuales de los modelos lineales pueden aplicarse aunque las hipótesis no se cumplan exactamente, bastará en muchos casos que lo hagan de una forma aproximada.

Por otra parte, una vez que hemos hecho la transformación, podemos hacer uso de los procedimientos vistos en el capítulo anterior, aunque, como es obvio, son procedimientos que sirven tan sólo para verificar transformaciones en la variable dependiente. En particular, pensamos que puede ser útil la aplicación iterativa -para los casos en que esto sea posible- del procedimiento de la R^2 (4.2). Podemos utilizarlo antes de efectuar la transformación y si no se cumple (pues de otra forma, lo que el teorema afirma es que no tendría sentido seguir transformando si las hipótesis básicas se satis

facen) aplicarlo después de ella y, de nuevo, si no se cumple se puede pensar en una nueva transformación y así sucesivamente.

A P E N D I C E

1. Tablas de funciones ortogonales utilizadas en el método de Box y Tidwell (2.2)

T A B L A

n	q=0.1	q=0.2	q=0.3	q=0.4	q=0.5
3	0.164 188	0.167 630	0.169 236	0.171 699	0.174 426
	-0.328 372	-0.335 258	-0.338 473	-0.343 395	-0.348 851
	0.164 186	0.167 628	0.169 238	0.171 696	0.174 426
Δ	0.001 617	0.006 743	0.015 466	0.028 300	0.045 636
4	0.223 211	0.227 277	0.230 371	0.235 360	0.242 019
	-0.226 206	-0.233 572	-0.239 360	-0.248 022	-0.258 367
	-0.217 217	-0.214 687	-0.212 395	-0.210 034	-0.209 322
	0.220 213	0.220 982	0.221 384	0.222 696	0.225 671
Δ	0.001 996	0.008 045	0.018 403	0.033 698	0.055 017
5	0.252 898	0.255 899	0.262 961	0.268 583	0.277 135
	-0.130 178	-0.135 714	-0.143 711	-0.151 149	-0.161 301
	-0.250 328	-0.249 611	-0.254 121	-0.255 869	-0.259 486
	-0.120 391	-0.117 229	-0.112 467	-0.109 146	-0.105 663
	0.248 007	0.246 656	0.247 338	0.247 582	0.249 316
Δ	0.002 195	0.008 831	0.020 538	0.037 385	0.060 869
6	0.271 279	0.277 440	0.281 119	0.289 587	0.298 854
	-0.058 219	-0.064 560	-0.069 628	-0.075 965	-0.084 156
	-0.216 925	-0.221 357	-0.223 704	-0.229 815	-0.235 597
	-0.211 782	-0.210 092	-0.207 228	-0.208 449	-0.208 227
	-0.049 203	-0.044 702	-0.041 514	-0.037 732	-0.034 404
	0.264 845	0.263 271	0.260 955	0.262 375	0.263 530
Δ	0.002 414	0.009 823	0.022 201	0.040 985	0.066 472
7	0.282 555	0.286 768	0.293 724	0.303 261	0.313 798
	-0.003 918	-0.008 251	-0.012 613	-0.018 351	-0.024 341
	-0.170 997	-0.174 834	-0.180 021	-0.187 409	-0.194 474
	-0.222 861	-0.222 596	-0.224 000	-0.227 183	-0.230 504
	-0.163 418	-0.159 647	-0.158 521	-0.155 314	-0.154 531
	0.003 670	0.007 065	0.009 847	0.013 999	0.017 388
	0.274 969	0.271 496	0.271 584	0.270 996	0.272 665
Δ	0.002 611	0.010 466	0.024 120	0.044 287	0.072 135

	0.290 020	0.295 276	0.804 083	0.312 805	0.324.454
	0.037 713	0.034 609	0.031 638	0.026 806	0.022 548
	-0.126 674	-0.131 162	-0.137 114	-0.143 951	-0.151 635
	-0.205 822	-0.207 994	-0.212 938	-0.215 563	-0.221 210
8	-0.202 252	-0.201 066	-0.202 056	-0.200 824	-0.202 133
	-0.118 340	-0.114 946	-0.112 345	-0.109 036	-0.106 501
	0.043 675	0.046 294	0.049 772	0.052 080	0.055 950
	0.281 681	0.278 990	0.278 960	0.277 683	0.278 527
Δ	0.002 801	0.011 298	0.026 222	0.047 647	0.077 653
	0.295 707	0.302 588	0.311 436	0.320 148	0.332 448
	0.070 473	0.068 566	0.066 226	0.063 172	0.060 111
	-0.087 358	-0.092 168	-0.098 555	-0.103 954	-0.111 649
	-0.179 584	-0.183 608	-0.188 723	-0.192 732	-0.199 407
9	-0.207 905	-0.209 318	-0.210 930	-0.212 182	-0.215 205
	-0.173 930	-0.172 504	-0.170 619	-0.170 062	-0.168 486
	-0.079 185	-0.076 076	-0.072 323	-0.070 066	-0.066 224
	0.074 885	0.077 309	0.080 113	0.082 449	0.085 296
	0.286 899	0.285 210	0.283 378	0.283 128	0.283 116
Δ	0.002 999	0.012 206	0.028 103	0.051 236	0.083 220

T A B L A

(continuación)

n	q=0,06	q=0.7	q=0.8	q=0.9
3	0.178 335	0.183 899	0.191 687	0.203 567
	-0.356 671	-0.367 800	-0.383 375	-0.407 135
	0.178 336	0.183 901	0.191 688	0.203 568
Δ	0.068 695	0.099 428	0.141 096	0.201 397
4	0.250 037	0.260 556	0.275 140	0.297 059
	-0.270 373	-0.285 711	-0.306 441	-0.336 866
	-0.209 364	-0.210 247	-0.212 538	-0.217 445
	-0.299 700	0.235 402	0.243 839	0.257 252
Δ	0.083 597	0.122 077	0.175 512	0.255 298
5	0.287 974	0.301 956	0.321 041	0.349 844
	-0.173 116	-0.187 834	-0.207 594	-0.236 974
	-0.264 880	-0.243 421	-0.281 630	-0.296 189
	-0.102 790	-0.100 336	-0.098 120	-0.096 078
	0.252 812	0.258 207	0.266 304	0.279 396
Δ	0.092 714	0.135 816	0.195 854	0.286 390

T A B L A (continuación)

n	q=0,6	q=0,7	q=0,8	q=0,9
6	0.311 428	0.327 596	0.349 781	0.383 380
	-0.094 444	-0.106 971	-0.123 977	-0.149 842
	-0.243 614	-0.253 812	-0.267 139	-0.286 300
	-0.208 948	-0.210 970	-0.214 564	-0.220 682
	-0.030 625	-0.026 717	-0.022 461	-0.017 269
	0.266 203	0.270 875	0.278 358	0.290 713
Δ	0.101 057	0.147 870	0.213 188	0.311 780
7	0.327 388	0.345 133	0.369 468	0.406 476
	-0.032 501	-0.042 656	-0.056 492	-0.078 177
	-0.203 361	-0.214 735	-0.229 491	-0.250 424
	-0.234 217	-0.239 548	-0.247 154	-0.258 528
	-0.153 321	-0.152 535	-0.152 571	-0.153 806
	0.020 912	0.025 138	0.030 221	0.036 920
0.275 100	0.279 204	0.286 019	0.297 539	
Δ	0.109 467	0.159 878	0.230 046	0.335 689
8	0.338 901	0.357 906	0.383 816	0.423 341
	0.016 667	0.008 877	-0.001 855	-0.019 355
	-0.160 507	-0.171 957	-0.186 860	-0.207 999
	-0.227 103	-0.234 645	-0.244 791	-0.259 405
	-0.203 837	-0.206 102	-0.209 723	-0.215 765
	-0.104 752	-0.102 839	-0.100 984	-0.099 352
0.059 650	0.063 746	0.069 019	0.076 358	
0.280 980	0.285 014	0.291 377	0.302 177	
Δ	0.117 899	0.172 064	0.247 041	0.359 407
9	0.347 522	0.367 475	0.394 686	0.436 178
	0.056 274	0.050 766	0.042 946	0.029 413
	-0.120 077	-0.130 956	-0.145 270	-0.165 731
	-0.206 641	-0.215 507	-0.227 158	-0.243 669
	-0.219 140	-0.223 877	-0.230 293	-0.239 845
	-0.168 623	-0.169 284	-0.170 568	-0.173 150
-0.063 569	-0.061 065	-0.058 341	-0.055 092	
0.089 150	0.093 594	0.098 861	0.106 433	
0.285 106	0.288 853	0.295 137	0.305 464	
Δ	0.126 283	0.184 195	0.264 151	0.383 239

2. En el procedimiento para obtener el sesgo de una transformación (4.1) se hace uso de las funciones enteras de segundo orden. Las definimos a continuación (ver p. ej. Ahlfors, 1953, pp. 38-42, 155);

Se dice que una función compleja f de variable compleja z es analítica en A si la derivada de f existe para todo $z \in A$.

Una función compleja de variable compleja es entera si es analítica en todo el plano.

Una función es entera de segundo orden si los radios de convergencia de las dos series

$$\sum_{n=0}^{\infty} \frac{1}{n!} f_0^{(2n)} z^n \quad \text{y} \quad \sum_{n=0}^{\infty} \frac{1}{n!} f_0^{(2n+1)} z^n$$

son infinito. Aquí $f_a^{(i)}$ representa la i -ésima derivada de f valuada en a .

B I B L I O G R A F I A

1. ANSCOMBE, F. J. (1948)
The transformation of Poisson, binomial and negative binomial data
Biometrika, vol. 35, pp. 246-54
2. BARTLETT, M. S. (1936)
The square root transformation in the Analysis of Variance
Journal of the Royal Statistical Society, suppl. vol. 3,
pp. 68-78
3. BARTLETT, M. S. (1947)
The use of transformations
Biometrics, vol. 3, pp. 39-52
4. BLISS, C. I. (1935)
The calculation of the dosage mortality curve
Annals of Applied Biology, vol. 22, pp. 134-67
5. BOX, G. E. P. y COX, D. R. (1964)
An analysis of transformations
Journal of the Royal Statistical Society, serie B, vol. 26
n. 2, pp. 211-52
6. BOX, G. E. P. y TIDWELL, P. W. (1962)
Transformation of the independent variable
Technometrics, vol. 4, n. 4, pp. 532-50
7. BOX, G. E. P. y WILSON, K. B. (1951)
On the attainment of optimum conditions
Journal of the Royal Statistical Society, serie B, vol. 13
pp. 1-38
8. BOX, M. J. (1971)
Bias in nonlinear estimation
Journal of the Royal Statistical Society, serie B, vol. 33
pp. 171-201

9. CURTISS, J. H. (1943)
On transformations used in the Analysis of Variance
Annals of Mathematical Statistics, vol. 14, n. 2, pp. 107-22
10. DOLBY, J. L. (1963)
A quick method for choosing a transformation
Technometrics, vol. 5, n. 3, pp. 317-25
11. DRAPER, N. R. y HUNTER, W. G. (1969)
Transformations: some examples revisited
Technometrics, vol. 11, n. 1, pp. 23-40
12. DRAPER, N. R. y SMITH, H. (1966)
Applied Regression Analysis
Wiley Intesciencé, USA
13. FINNEY, D. J. (1964)
Probit Analysis
Cambridge University Press, Cambridge
14. FISHER, R. A. (1925)
Statistical Methods of Research Workers
Primera edición, Griffin, Londres
15. FISHER, R. A. y YATES, F. (1957)
Statistical tables for biological, agricultural and medical
research
Quinta edición, Oliver and Boyd, Edimburgo
16. GRANGER, C. W. J. y NEWBOLD, P. (1976)
The use of R^2 to determinate the appropriate transformation
of Regression variables
Journal of Econometrics, vol. 4, pp. 205-10
17. GRAYBILL, Franklin A. (1961)
An introduction to linear statistical models
Mc Graw Hill series in Probability and Statistics, vol. 1
18. GUERRERO, G. V. M. y HERNANDEZ, A. F. M. (1974)
El uso de transformaciones en los modelos lineales
Tesis de Actuaría, U N A M

19. HILL, W. J. (1966)
Statistical techniques for Model Building
Ph. D. Thesis, University of Wisconsin
20. HOYLE, M. H. (1973)
Transformations - An Introduction and a Bibliography
International Statistical Review, vol. 41, n. 2, pp. 203-23
21. JOHNSTON, J. (1972)
Econometric Methods
2a. edición, Mc Graw-Hill Kogakusha, Tokyo
22. KIHLEBERG, Herson y Shotz (1967)
Square root transformations revisited
Congreso anual de la American Statistical Association
23. KOWALSKI, C. J. y Tarter, M. E. (1969)
Coordinate transformations to Normality and the power of
normal tests for independence
Biometrika, vol. 56, pp. 139-48
24. LEVENE, H. (1960)
Robust tests for equality of variances
Contributions to Probability and Statistics, Stanford Uni-
versity Press
25. LINDSEY, J. K. (1975)
The roles of Transformations to Normality
Biometrics, vol. 31, n. 1, pp. 247-9
26. MERRINGTON, M. (1941)
Numerical aproximations to percentage points of the
chi-square distribution
Biometrika, vol. 32, pp: 200-2
27. NEYMAN, J. y SCOTT, E. L. (1960)
Correction for bias introduced by a transformation of
variables
Annals of the Mathematical Statistics, vol. 31, pp. 643-55

28. PEARSON, E. S. (1938)
The probability integral transformation for testing goodness of fit and combining independent tests of significance
Biometrika, vol. 30, pp. 134-48
29. TUKEY, J. W. (1957)
On the comparative anatomy of transformations.
Annals of Mathematical Statistics, vol. 28, pp. 602-32
30. TURNER, M. E. MONROE, R. J. y LUCAS, H. L (1959)
The single process law: a study in nonlinear regression
Institute of Statistics, Raleigh, N. C. mimeo series, n. 235
31. TURNER, M.E., MONROE, R.J. y LUCAS, H. L. (1961)
Regression and nonlinear Path Analysis
Biometrics, vol. 17, pp. 120-43
32. WILSON, E. V. y HILFERTY, M. M. (1931)
The distribution of chi-square
Proc. Nat. Acad. Sa, vol. 17, pp. 684-8
33. WONNACOTT, R. J. y WONNACOTT, T. H. (1970)
Econometrics
Wiley International Edition